



AD

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7: C12N 15/12, C07K 14/72, G01N 33/50, 33/566

A1

(11) International Publication Number:

WO 00/22129

(43) International Publication Date:

20 April 2000 (20.04.00)

(21) International Application Number:

PCT/US99/23938

(22) International Filing Date:

12 October 1999 (12.10.99)

(30) Priority Data:

09/170,496

13 October 1998 (13.10.98)

US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application

US

09/170,496 (CIP)

Filed on

13 October 1998 (13.10.98)

(71) Applicant (for all designated States except US): ARENA PHARMACELTICALS, INC. [US/US]; 6166 Nancy Ridge Drive, San Diego, CA 92121 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): BEHAN, Dominic, P. [GB/US]; 11472 Roxboro Court, San Diego, CA 92131 (US). CHALMERS, Derek, T. [GB/US]; 347 Longden Lane, Solana Beach, CA 92075 (US). LIAW, Chen, W. [US/US]; 7668 Salix Place, San Diego, CA 92129 (US).

(74) Agents: MILLER, Suzanne, E. et al.; Woodcock Washburn Kurtz Mackiewicz & Norris LLP, 46th floor, One Liberty Place, Philadelphia, PA 19103 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: NON-ENDOGENOUS, CONSTITUTIVELY ACTIVATED HUMAN G PROTEIN-COUPLED RECEPTORS

(57) Abstract

Disclosed herein are constitutively activated, non-endogenous versions of endogenous human G protein-coupled receptors comprising (a) the following amino acid sequence region (C-terminus to N-terminus orientation) and/or (b) the following nucleic acid sequence region (3' to 5' orientation) transversing the transmembrane-6 (TM6) and intracellular loop-3 (IC3) regions of the GPCR: (a) P¹ AA₁₅ X and/or (b) Pcodon (AA-codon)₁₅ X_{codon}, respectively. In a most preferred embodiment, P¹ and Pcodon are endogenous proline and an endogenous nucleic acid encoding region encoding proline, respectively, located within TM6 of the non-endogenous GPCR; AA₁₅ and (AA-codon)₁₅ are 15 endogenous amino acid residues and 15 codons encoding endogenous amino acid residues, respectively; and X and X_{codon} are non-endogenous lysine and a non-endogenous nucleic acid encoding region encoding lysine, respectively, located within IC3 of the non-endogenous GPCR. Because it is most preferred that the non-endogenous human GPCRs which incorporate these mutations are incorporated into manimalian cells and utilized for the screening of the candidate compounds, the non-endogenous human GPCR incorporating the mutation need not be purified and isolated *per se* (i.e., these are incorporated within the cellular membrane of a mammalian cell), although such purified and isolated non-endogenous human GPCRs are well within the purview of this disclosure.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia	
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia	
ΥA	Austria	FR	France	LU	Luxembourg	SN	Senegal	
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland	
AZ.	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad	
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo	
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan	
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan	
BF	Burkina Paso	GR	Greece		Republic of Macedonia	TR	Turkey	
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago	
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine	
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda	
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America	
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan	
CF	Central African Republic	JP	Japan	NE	Niger	VN	Vict Nam	
CG	Congo	KE	Кепуа	NL	Netherlands	YU	Yugoslavia	
СН	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe	
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand			
CM	Cameroon		Republic of Korea	PL	Poland			
CN	China	KR	Republic of Korea	PT	Portugal			
CU	Cuba	KZ	Kazakstan	RO	Romania			
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation			
DE	Germany	LI	Liechtenstein	SD	Sudan			
DK	Denmark	LK	Sri Lanka	SE	Sweden			
EE	Estonia	LR	Liberia	SG	Singapore			
l								

NON-ENDOGENOUS, CONSTITUTIVELY ACTIVATED HUMAN G PROTEIN-COUPLED RECEPTORS

The benefits of commonly owned U.S. Serial Number 09/170,496, filed October 13, 1998, U.S. Serial Number 08/839, 449 filed April 14, 1997 (now abandoned), U.S. Serial Number 09/060,188, filed April 14, 1998; U.S. Provisional Number 60/090,783, filed June 26, 1998; and U.S. Provisional Number 60/095,677, filed on August 7, 1998, are hereby claimed. Each of the foregoing applications are incorporated by reference herein in their entirety.

FIELD OF THE INVENTION

The invention disclosed in this patent document relates to transmembrane receptors, and more particularly to human G protein-coupled receptors (GPCRs) which have been altered such that altered GPCRs are constitutively activated. Most preferably, the altered human GPCRs are used for the screening of therapeutic compounds.

BACKGROUND OF THE INVENTION

Although a number of receptor classes exist in humans, by far the most abundant and therapeutically relevant is represented by the G protein-coupled receptor (GPCR or GPCRs) class. It is estimated that there are some 100,000 genes within the human genome, and of these, approximately 2% or 2,000 genes, are estimated to code for GPCRs. Of these, there are approximately 100 GPCRs for which the endogenous ligand that binds to the GPCR has been identified. Because of the significant time-lag that exists between the discovery of an endogenous GPCR and its endogenous ligand, it can be presumed that the remaining 1,900 GPCRs will be identified and characterized long before the endogenous ligands for these receptors are identified. Indeed, the rapidity by which the Human Genome Project is sequencing the 100,000 human genes indicates that the remaining human GPCRs will be fully sequenced within the next few years. Nevertheless, and despite the efforts to sequence the human genome, it is still very unclear as to how scientists will be able to rapidly, effectively and efficiently exploit this information to improve and enhance the human condition. The present invention is geared towards this important objective.

Receptors, including GPCRs, for which the endogenous ligand has been identified are referred to as "known" receptors, while receptors for which the endogenous ligand has not been identified are referred to as "orphan" receptors. This distinction is not merely semantic, particularly in the case of GPCRs. GPCRs represent an important area for the development of pharmaceutical products: from approximately 20 of the 100 known GPCRs, 60% of all prescription pharmaceuticals have been developed. Thus, the orphan GPCRs are to the pharmaceutical industry what gold was to California in the late 19th century – an opportunity to drive growth, expansion, enhancement and development. A serious drawback exists, however,

15

5

10

15

20

with orphan receptors relative to the discovery of novel therapeutics. This is because the traditional approach to the discovery and development of pharmaceuticals has required access to both the receptor *and* its endogenous ligand. Thus, heretofore, orphan GPCRs have presented the art with a tantalizing and undeveloped resource for the discovery of pharmaceuticals.

Under the traditional approach to the discovery of potential therapeutics, it is generally the case that the receptor is first identified. Before drug discovery efforts can be initiated, elaborate, time consuming and expensive procedures are typically put into place in order to identify, isolate and generate the receptor's endogenous ligand - this process can require from between 3 and ten years per receptor, at a cost of about \$5million (U.S.) per receptor. These time and financial resources must be expended before the traditional approach to drug discovery can commence. This is because traditional drug discovery techniques rely upon so-called "competitive binding assays" whereby putative therapeutic agents are "screened" against the receptor in an effort to discover compounds that either block the endogenous ligand from binding to the receptor-("antagonists"), or enhance or mimic the effects of the ligand binding to the receptor ("agonists"). The overall objective is to identify compounds that prevent cellular activation when the ligand binds to the receptor (the antagonists), or that enhance or increase cellular activity that would otherwise occur if the ligand was properly binding with the receptor (the agonists). Because the endogenous ligands for orphan GPCRs are by definition not identified, the ability to discover novel and unique therapeutics to these receptors using traditional drug discovery techniques is not possible. The present invention, as will be set forth in greater detail below, overcomes these and other severe limitations created by such traditional drug discovery techniques.

GPCRs share a common structural motif. All these receptors have seven sequences of between 22 to 24 hydrophobic amino acids that form seven alpha helices, each of which spans the

membrane (each span is identified by number, *i.e.*, transmembrane-1 (TM-1), transmebrane-2 (TM-2), etc.). The transmembrane helices are joined by strands of amino acids between transmembrane-2 and transmembrane-3, transmembrane-4 and transmembrane-5, and transmembrane-6 and transmembrane-7 on the exterior, or "extracellular" side, of the cell membrane (these are referred to as "extracellular" regions 1, 2 and 3 (EC-1, EC-2 and EC-3), respectively). The transmembrane helices are also joined by strands of amino acids between transmembrane-1 and transmembrane-2, transmembrane-3 and transmembrane-4, and transmembrane-5 and transmembrane-6 on the interior, or "intracellular" side, of the cell membrane (these are referred to as "intracellular" regions 1, 2 and 3 (IC-1, IC-2 and IC-3), respectively). The "carboxy" ("C") terminus of the receptor lies in the intracellular space within the cell, and the "amino" ("N") terminus of the receptor lies in the extracellular space outside of the cell. The general structure of G protein-coupled receptors is depicted in Figure 1.

Generally, when an endogenous ligand binds with the receptor (often referred to as "activation" of the receptor), there is a change in the conformation of the intracellular region that allows for coupling between the intracellular region and an intracellular "G-protein." Although other G proteins exist, currently, Gq, Gs, Gi, and Go are G proteins that have been identified. Endogenous ligand-activated GPCR coupling with the G-protein begins a signaling cascade process (referred to as "signal transduction"). Under normal conditions, signal transduction ultimately results in cellular activation or cellular inhibition. It is thought that the IC-3 loop as well as the carboxy terminus of the receptor interact with the G protein. A principal focus of this invention is directed to the transmembrane-6 (TM6) region and the intracellular-3 (IC3) region of the GPCR.

Under physiological conditions, GPCRs exist in the cell membrane in equilibrium between

PCT/US99/23938

5

15

20

two different conformations: an "inactive" state and an "active" state. As shown schematically in Figure 2, a receptor in an inactive state is unable to link to the intracellular signaling transduction pathway to produce a biological response. Changing the receptor conformation to the active state allows linkage to the transduction pathway (via the G-protein) and produces a biological response.

A receptor may be stabilized in an active state by an endogenous ligand or a compound such as a drug. Recent discoveries, including but not exclusively limited to modifications to the amino acid sequence of the receptor, provide means other than endogenous ligands or drugs to promote and stabilize the receptor in the active state conformation. These means effectively stabilize the receptor in an active state by simulating the effect of an endogenous ligand binding to the receptor. Stabilization by such ligand-independent means is termed "constitutive receptor activation."

As noted above, the use of an orphan receptor for screening purposes has not been possible. This is because the traditional "dogma" regarding screening of compounds mandates that the ligand for the receptor be known. By definition, then, this approach has no applicability with respect to orphan receptors. Thus, by adhering to this dogmatic approach to the discovery of therapeutics, the art, in essence, has taught and has been taught to forsake the use of orphan receptors unless and until the endogenous ligand for the receptor is discovered. Given that there are an estimated 2,000 G protein coupled receptors, the majority of which are orphan receptors, such dogma castigates a creative, unique and distinct approach to the discovery of therapeutics.

Information regarding the nucleic acid and/or amino acid sequences of a variety of GPCRs is summarized below in Table A. Because an important focus of the invention disclosed herein is directed towards orphan GPCRs, many of the below-cited references are related to orphan GPCRs. However, this list is not intended to imply, nor is this list to be construed, legally or

otherwise, that the invention disclosed herein is only applicable to orphan GPCRs or the specific GPCRs listed below. Additionally, certain receptors that have been isolated are not the subject of publications per se; for example, reference is made to a G Protein-Coupled Receptor database on the "world-wide web" (neither the named inventors nor the assignee have any affiliation with this site) that lists GPCRs. Other GPCRs are the subject of patent applications owned by the present assignee and these are not listed below (including GPR3, GPR6 and GPR12; see U.S. Provisional Number 60/094879):

Table A

	Receptor Name	Publication Reference
10	GPR1	23 Genomics 609 (1994)
	GPR4	14 DNA and Cell Biology 25 (1995)
	GPR5	14 DNA and Cell Biology 25 (1995)
	GPR7	28 Genomics 84 (1995)
	GPR8	28 Genomics 84 (1995)
15	GPR9	184 J. Exp. Med. 963 (1996)
	GPR10	29 Genomics 335 (1995)
	GPR15	32 Genomics 462 (1996)
	GPR17	70 J Neurochem. 1357 (1998)
	GPR18	42 Genomics 462 (1997)
20	GPR20	187 Gene 75 (1997)
•	GPR21	187 Gene 75 (1997)
	GPR22	187 Gene 75 (1997)
•	GPR24	398 FEBS Lett. 253 (1996)
	GPR30	45 Genomics 607 (1997)
25	GPR31	42 Genomics 519 (1997)
1	GPR32	50 Genomics 281 (1997)
	GPR40	239 Biochem. Biophys.
		Res. Commun. 543 (1997)
	GPR41	239 Biochem. Biophys.
		Res. Commun. 543 (1997)
	GPR43	239 Biochem. Biophys.
•		Res. Commun. 543 (1997)
30	APJ	136 Gene 355 (1993)
	BLR1	22 Eur. J. Immunol. 2759 (1992)
	CEPR	231 Biochem. Biophys.
		Res. Commun. 651 (1997)
•	EBI1	23 Genomics 643 (1994)
	EBI2	67 J. Virol. 2209 (1993)
35	ETBR-LP2	424 FEBS Lett. 193 (1998)
	GPCR-CNS	54 Brain Res. Mol. Brain Res. 152 (1998);
		45 Genomics 68 (1997)
	GPR-NGA	394 FEBS Lett. 325 (1996)
	Н9	386 FEBS Lett 219 (1996)

BNSDOCID <WO 002212941 I

-7-

HBA954	1261 Biochim. Biophys. Acta 121 (1995)		
HG38	247 Biochem. Biophys.		
	Res. Commun. 266 (1998)		
HM74	5 Int. Immunol. 1239 (1993)		
OGR1 35 Genomics 397 (1996)			
V28	163 Gene 295 (1995)		

As will be set forth and disclosed in greater detail below, utilization of a mutational cassette to modify the endogenous sequence of a human GPCR leads to a constitutively activated version of the human GPCR. These non-endogenous, constitutively activated versions of human GPCRs can be utilized, *inter alia*, for the screening of candidate compounds to directly identify compounds of, *e.g.*, therapeutic relevance.

SUMMARY OF THE INVENTION

Disclosed herein is a non-endogenous, human G protein-coupled receptor comprising

(a) as a most preferred amino acid sequence region (C-terminus to N-terminus orientation)

and/or (b) as a most preferred nucleic acid sequence region (3' to 5' orientation) transversing

the transmembrane-6 (TM6) and intracellular loop-3 (IC3) regions of the GPCR:

(a)
$$P^1 AA_{15} X$$

wherein:

10

15

- (1) P¹ is an amino acid residue located within the TM6 region of the GPCR, where P¹ is selected from the group consisting of (i) the endogenous GPCR's proline residue, and (ii) a non-endogenous amino acid residue other than proline;
- (2) AA₁₅ are 15 amino acids selected from the group consisting of

(a) the endogenous GPCR's amino acids (b) non-endogenous amino acid residues, and (c) a combination of the endogenous GPCR's amino acids and non-endogenous amino acids, excepting that none of the 15 endogenous amino acid residues that are positioned within the TM6 region of the GPCR is proline; and

X is a non-endogenous amino acid residue located within the (3) IC3 region of said GPCR, preferably selected from the group consisting of lysine, hisitidine and arginine, and most preferably lysine, excepting that when the endogenous amino acid at position X is lysine, then X is an amino acid other than lysine, preferably alanine;

and/or

5

10

(b) Pcodon (AA-codon)15 Xcodon

15 wherein:

> P^{codon} is a nucleic acid sequence within the TM6 region of the (1) GPCR, where P^{codon} encodes an amino acid selected from the group consisting of (i) the endogenous GPCR's proline residue, and (ii) a non-endogenous amino acid residue other than proline;

20

(AA-codon)₁₅ are 15 codons encoding 15 amino acids selected (2) from the group consisting of (a) the endogenous GPCR's amino acids (b) non-endogenous amino acid residues and (c) a combination of the endogenous GPCR's amino acids and non-

PCT/US99/23938

10

15

20

endogenous amino acids, excepting that none of the 15 endogenous codons within the TM6 region of the GPCR encodes a proline amino acid residue; and

(3) X_{codon} is a nucleic acid encoding region residue located within the IC3 region of said GPCR, where X_{codon} encodes a non-endogenous amino acid, preferably selected from the group consisting of lysine, hisitidine and arginine, and most preferably lysine, excepting that when the endogenous encoding region at position X_{codon} encodes the amino acid lysine, then X_{codon} encodes an amino acid other than lysine, preferably alanine.

The terms endogenous and non-endogenous in reference to these sequence cassettes are relative to the endogenous GPCR. For example, once the endogenous proline residue is located within the TM6 region of a particular GPCR, and the 16th amino acid therefrom is identified for mutation to constitutively activate the receptor, it is also possible to mutate the endogenous proline residue (*i.e.*, once the marker is located and the 16th amino acid to be mutated is identified, one may mutate the marker itself), although it is most preferred that the proline residue not be mutated. Similarly, and while it is most preferred that AA₁₅ be maintained in their endogenous forms, these amino acids may also be mutated. The only amino acid that must be mutated in the non-endogenous version of the human GPCR is X *i.e.*, the endogenous amino acid that is 16 residues from P¹ cannot be maintained in its endogenous form and must be mutated, as further disclosed herein. Stated again, while it is preferred that in the non-endogenous version of the human GPCR, P¹ and AA₁₅ remain in their endogenous forms (*i.e.*, identical to their wild-type forms), once X is identified and mutated, any and/or all of P¹ and AA₁₅ can be mutated. This applies to the nucleic

acid sequences as well. In those cases where the endogenous amino acid at position X is lysine, then in the non-endogenous version of such GPCR, X is an amino acid other than lysine, preferably alanine.

Accordingly, and as a hypothetical example, if the endogenous GPCR has the following endogenous amino acid sequence at the above-noted positions:

P-AACCTTGGRRRDDDE -O

then any of the following exemplary and hypothetical cassettes would fall within the scope of the disclosure (non-endogenous amino acids are set forth in bold):

P-AACCTTGGRRRDDDE -K

10

P-AACCTTHIGRRDDDE -K

P-ADEETTGGRRRDDDE -A

P-LLKFMSTWZLVAAPO -K

A-LLKFMSTWZLVAAPO-K

It is also possible to add amino acid residues within AA₁₅, but such an approach is not particularly advanced. Indeed, in the most preferred embodiments, the only amino acid that differs in the non-endogenous version of the human GPCR as compared with the endogenous version of that GPCR is the amino acid in position X; mutation of this amino acid itself leads to constitutive activation of the receptor.

Thus, in particularly preferred embodiments, P^1 and P^{codon} are endogenous proline and an endogenous nucleic acid encoding region encoding proline, respectively; and X and X_{codon} are non-endogenous lysine or alanine and a non-endogenous nucleic acid encoding region encoding lysine or alanine, respectively, with lysine being most preferred. Because it is most preferred that the non-endogenous versions of the human GPCRs which incorporate these mutations are

ביו יאשריניניט עוטטינישריי וי

15

incorporated into mammalian cells and utilized for the screening of candidate compounds, the non-endogenous human GPCR incorporating the mutation need not be purified and isolated *per se* (*i.e.*, these are incorporated within the cellular membrane of a mammalian cell), although such purified and isolated non-endogenous human GPCRs are well within the purview of this disclosure. Genetargeted and transgenic non-human mammals (preferably rats and mice) incorporating the non-endogenous human GPCRs are also within the purview of this invention; in particular, genetargeted mammals are most preferred in that these animals will incorporate the non-endogenous versions of the human GPCRs in place of the non-human mammal's endogenous GPCR-encoding region (techniques for generating such non-human mammals to replace the non-human mammal's protein encoding region with a human encoding region are well known; see, for example, U.S. Patent No. 5,777,194.)

It has been discovered that these changes to an endogenous human GPCR render the GPCR constitutively active such that, as will be further disclosed herein, the non-endogenous, constitutively activated version of the human GPCR can be utilized for, *inter alia*, the direct screening of candidate compounds without the need for the endogenous ligand. Thus, methods for using these materials, and products identified by these methods are also within the purview of the following disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a generalized structure of a G protein-coupled receptor with the numbers

assigned to the transmembrane helixes, the intracellular loops, and the extracellular loops.

Figure 2 schematically shows the two states, active and inactive, for a typical G protein coupled receptor and the linkage of the active state to the second messenger transduction pathway.

Figure 3 is a sequence diagram of the preferred vector pCMV, including restriction enzymen site locations.

Figure 4 is a diagrammatic representation of the signal measured comparing pCMV, non-endogenous, constitutively active GPR30 inhibition of GPR6-mediated activation of CRE-Luc reporter with endogenous GPR30 inhibition of GPR6-mediated activation of CRE-Luc reporter.

Figure 5 is a diagrammatic representation of the signal measured comparing pCMV, nonendogenous, constitutively activated GPR17 inhibition of GPR3-mediated activation of CRELuc reporter with endogenous GPR17 inhibition of GPR3-mediated activation of CRE-Luc
reporter.

Figure 6 provides diagrammatic results of the signal measured comparing control pCMV, endogenous APJ and non-endogenous APJ.

Figure 7 provides an illustration of IP₃ production from non-endogenous human 5-HT_{2A} receptor as compared to the endogenous version of this receptor.

Figure 8 are dot-blot format results for GPR1 (8A), GPR30 (8B) and APJ (8C).

DETAILED DESCRIPTION

The scientific literature that has evolved around receptors has adopted a number of terms to refer to ligands having various effects on receptors. For clarity and consistency, the following definitions will be used throughout this patent document. To the extent that these definitions conflict with other definitions for these terms, the following definitions shall control:

AGONISTS shall mean compounds that activate the intracellular response when they bind to the receptor, or enhance GTP binding to membranes.

PCT/US99/23938

25

30

- 13 -

AMINO ACID ABBREVIATIONS used herein are set below:

		ALANINE	ALA	Α
		ARGININE	ARG	R
		ASPARAGINE	ASN	N
5		ASPARTIC ACID	ASP	D
	* 2 5	CYSTEINE	CYS	С
٠,	-:/	GLUTAMIC ACID	GLU	Ε.
	-	GLUTAMINE	GLN	Q
		GLYCINE	GLY	G
10		HISTIDINE	HIS	Н
•		ISOLEUCINE	ILE	1
		LEUCINE	LEU	L
		LYSINE	LYS	K
		METHIONINE	MET	M
15		PHENYLALANINE	PHE	F
		PROLINE	PRO	P
		SERINE	SER	S
	ć.;	THREONINE	THR	T
		TRYPTOPHAN	TRP	w
20		TYROSINE	TYR	Y
		VALINE	VAL	V

PARTIAL AGONISTS shall mean compounds which activate the intracellular response when they bind to the receptor to a lesser degree/extent than do agonists, or enhance GTP binding to membranes to a lesser degree/extent than do agonists

ANTAGONIST shall mean compounds that competitively bind to the receptor at the same site as the agonists but which do not activate the intracellular response initiated by the active form of the receptor, and can thereby inhibit the intracellular responses by agonists or partial agonists. ANTAGONISTS do not diminish the baseline intracellular response in the absence of an agonist or partial agonist.

CANDIDATE COMPOUND shall mean a molecule (for example, and not limitation, a chemical compound) which is amenable to a screening technique. Preferably, the phrase

"candidate compound" does not include compounds which were publicly known to be compounds selected from the group consisting of inverse agonist, agonist or antagonist to a receptor, as previously determined by an indirect identification process ("indirectly identified compound"); more preferably, not including an indirectly identified compound which has previously been determined to have therapeutic efficacy in at least one mammal; and, most preferably, not including an indirectly identified compound which has previously been determined to have therapeutic utility in humans.

CODON shall mean a grouping of three nucleotides (or equivalents to nucleotides) which generally comprise a nucleoside (adenosine (A), guanosine (G), cytidine (C), uridine (U) and thymidine (T)) coupled to a phosphate group and which, when translated, encodes an amino acid.

COMPOUND EFFICACY shall mean a measurement of the ability of a compound to inhibit or stimulate receptor functionality, as opposed to receptor binding affinity. A preferred means of detecting compound efficacy is via measurement of, e.g., [35S]GTPyS binding, as further disclosed in the Example section of this patent document.

CONSTITUTIVELY ACTIVATED RECEPTOR shall mean a receptor subject to constitutive receptor activation. In accordance with the invention disclosed herein, a non-endogenous, human constitutively activated G protein-coupled receptor is one that has been mutated to include the amino acid cassette P¹AA₁₅X, as set forth in greater detail below.

in the active state by means other than binding of the receptor with its endogenous ligand or a chemical equivalent thereof. Preferably, a G protein-coupled receptor subjected to constitutive receptor activation in accordance with the invention disclosed herein evidences at least a 10% difference in response (increase or decrease, as the case may be) to the signal measured for

ENZUCCIU -MU UUSSISCAII

10

constitutive activation as compared with the endogenous form of that GPCR, more preferably, about a 25% difference in such comparative response, and most preferably about a 50% difference in such comparative response. When used for the purposes of directly identifying candidate compounds, it is most preferred that the signal difference be at least about 50% such that there is a sufficient difference between the endogenous signal and the non-endogenous signal to differentiate between selected candidate compounds. In most instances, the "difference" will be an increase in signal; however, with respect to Gs-coupled GPCRS, the "difference" measured is preferably a decrease, as will be set forth in greater detail below.

CONTACT or CONTACTING shall mean bringing at least two moieties together,

whether in an in vitro system or an in vivo system.

DIRECTLY IDENTIFYING or DIRECTLY IDENTIFIED, in relationship to the phrase "candidate compound", shall mean the screening of a candidate compound against a constitutively activated G protein-coupled receptor, and assessing the compound efficacy of such compound. This phrase is, under no circumstances, to be interpreted or understood to be encompassed by or to encompass the phrase "indirectly identifying" or "indirectly identified."

15

20

ENDOGENOUS shall mean a material that is naturally produced by the genome of the species. ENDOGENOUS in reference to, for example and not limitation, GPCR, shall mean that which is naturally produced by a human, an insect, a plant, a bacterium, or a virus. By contrast, the term NON-ENDOGENOUS in this context shall mean that which is not naturally produced by the genome of a species. For example, and not limitation, a receptor which is not constitutively active in its endogenous form, but when mutated by using the cassettes disclosed herein and thereafter becomes constitutively active, is most preferably referred to herein as a "non-endogenous, constitutively activated receptor." Both terms can be utilized to describe both "in

vivo" and "in vitro" systems. For example, and not limitation, in a screening approach, the endogenous or non-endogenous receptor may be in reference to an in vitro screening system whereby the receptor is expressed on the cell-surface of a mammalian cell. As a further example and not limitation, where the genome of a mammal has been manipulated to include a non-endogenous constitutively activated receptor, screening of a candidate compound by means of an in vivo system is viable.

therein. In the case of a prokaryotic Host Cell, a Plasmid is typically replicated as an autonomous molecule as the Host Cell replicates (generally, the Plasmid is thereafter isolated for introduction into a eukaryotic Host Cell); in the case of a eukaryotic Host Cell, a Plasmid is integrated into the cellular DNA of the Host Cell such that when the eukaryotic Host Cell replicates, the Plasmid replicates. Preferably, for the purposes of the invention disclosed herein, the Host Cell is eukaryotic, more preferably, mammalian, and most preferably selected from the group consisting of 293, 293T and COS-7 cells.

INDIRECTLY IDENTIFYING or INDIRECTLY IDENTIFIED means the traditional approach to the drug discovery process involving identification of an endogenous ligand specific for an endogenous receptor, screening of candidate compounds against the receptor for determination of those which interfere and/or compete with the ligand-receptor interaction, and assessing the efficacy of the compound for affecting at least one second messenger pathway associated with the activated receptor.

INHIBIT or INHIBITING, in relationship to the term "response" shall mean that a response is decreased or prevented in the presence of a compound as opposed to in the absence of the compound.

5

15

INVERSE AGONISTS shall mean compounds which bind to either the endogenous form of the receptor or to the constitutively activated form of the receptor, and which inhibit the baseline intracellular response initiated by the active form of the receptor below the normal base level of activity which is observed in the absence of agonists or partial agonists, or decrease GTP binding to membranes. Preferably, the baseline intracellular response is inhibited in the presence of the inverse agonist by at least 30%, more preferably by at least 50%, and most preferably by at least 75%, as compared with the baseline response in the absence of the inverse agonist.

KNOWN RECEPTOR shall mean an endogenous receptor for which the endogenous ligand specific for that receptor has been identified.

10 LIGAND shall mean an endogenous, naturally occurring molecule specific for an endogenous, naturally occurring receptor.

MUTANT or MUTATION in reference to an endogenous receptor's nucleic acid and/or amino acid sequence shall mean a specified change or changes to such endogenous sequences such that a mutated form of an endogenous, non-constitutively activated receptor evidences constitutive activation of the receptor. In terms of equivalents to specific sequences, a subsequent mutated form of a human receptor is considered to be equivalent to a first mutation of the human receptor if (a) the level of constitutive activation of the subsequent mutated form of the receptor is substantially the same as that evidenced by the first mutation of the receptor; and (b) the percent sequence (amino acid and/or nucleic acid) homology between the subsequent mutated form of the receptor and the first mutation of the receptor is at least about 80%, more preferably at least about 90% and most preferably at least 95%. Ideally, and owing to the fact that the most preferred cassettes disclosed herein for achieving constitutive activation includes a single amino acid and/or codon change between the endogenous and the non-endogenous forms of the GPCR (i.e. X or

 X_{codon}), the percent sequence homology should be at least 98%.

ORPHAN RECEPTOR shall mean an endogenous receptor for which the endogenous ligand specific for that receptor has not been identified or is not known.

PHARMACEUTICAL COMPOSITION shall mean a composition comprising at least one active ingredient, whereby the composition is amenable to investigation for a specified, efficacious outcome in a mammal (for example, and not limitation, a human). Those of ordinary skill in the art will understand and appreciate the techniques appropriate for determining whether an active ingredient has a desired efficacious outcome based upon the needs of the artisan.

PLASMID shall mean the combination of a Vector and cDNA. Generally, a Plasmid is introduced into a Host Cell for the purpose of replication and/or expression of the cDNA as a protein.

STIMULATE or STIMULATING, in relationship to the term "response" shall mean that a response is increased in the presence of a compound as opposed to in the absence of the compound.

TRANSVERSE or TRANSVERSING, in reference to either a defined nucleic acid sequence or a defined amino acid sequence, shall mean that the sequence is located within at least two different and defined regions. For example, in an amino acid sequence that is 10 amino acid moieties in length, where 3 of the 10 moieties are in the TM6 region of a GPCR and the remaining 7 moieties are in the IC3 region of the GPCR, the 10 amino acid moiety can be described as transversing the TM6 and IC3 regions of the GPCR.

VECTOR in reference to cDNA shall mean a circular DNA capable of incorporating at least one cDNA and capable of incorporation into a Host Cell.

The order of the following sections is set forth for presentational efficiency and is not

15

- 19 -

intended, nor should be construed, as a limitation on the disclosure or the claims to follow.

A. Introduction

20

The traditional study of receptors has always proceeded from the a priori assumption (historically based) that the endogenous ligand must first be identified before discovery could proceed to find antagonists and other molecules that could affect the receptor. Even in cases where an antagonist might have been known first, the search immediately extended to looking for the endogenous ligand. This mode of thinking has persisted in receptor research even after the discovery of constitutively activated receptors. What has not been heretofore recognized is that it is the active state of the receptor that is most useful for discovering agonists, partial agonists, and inverse agonists of the receptor. For those diseases which result from an overly active receptor or an under-active receptor, what is desired in a therapeutic drug is a compound which acts to diminish the active state of a receptor or enhance the activity of the receptor, respectively, not necessarily a drug which is an antagonist to the endogenous ligand. This is because a compound that reduces or enhances the activity of the active receptor state need not bind at the same site as the endogenous ligand. Thus, as taught by a method of this invention, any search for therapeutic compounds should start by screening compounds against the ligand-independent active state.

Screening candidate compounds against non-endogenous, constitutively activated GPCRs allows for the direct identification of candidate compounds which act at these cell surface receptors, without requiring any prior knowledge or use of the receptor's endogenous ligand. By determining areas within the body where the endogenous version of such GPCRs are expressed and/or over-expressed, it is possible to determine related disease/disorder states which are associated with the expression and/or over-expression of these receptors; such an approach is disclosed in this patent document.

B. Disease/Disorder Identification and/or Selection

Most preferably, inverse agonists to the non-endogenous, constitutively activated GPCRs can be identified using the materials of this invention. Such inverse agonists are ideal candidates as lead compounds in drug discovery programs for treating diseases related to these receptors. Because of the ability to directly identify inverse agonists, partial agonists or agonists to these receptors, thereby allowing for the development of pharmaceutical compositions, a search, for diseases and disorders associated with these receptors is possible. For example, scanning both diseased and normal tissue samples for the presence of these receptor now becomes more than an academic exercise or one which might be pursued along the path of identifying, in the case of an orphan receptor, an endogenous ligand. Tissue scans can be conducted across a broad range of healthy and diseased tissues. Such tissue scans provide a preferred first step in associating a specific receptor with a disease and/or disorder.

Preferably, the DNA sequence of the endogenous GPCR is used to make a probe for either radiolabeled cDNA or RT-PCR identification of the expression of the GPCR in tissue samples.

The presence of a receptor in a diseased tissue, or the presence of the receptor at elevated or decreased concentrations in diseased tissue compared to a normal tissue, can be preferably utilized to identify a correlation with that disease. Receptors can equally well be localized to regions of organs by this technique. Based on the known functions of the specific tissues to which the receptor is localized, the putative functional role of the receptor can be deduced.

20 C. A "Human GPCR Proline Marker" Algorithm and the Creation of Non-Endogenous, Constitutively-Active Human GPCRs

Among the many challenges facing the biotechnology arts is the unpredictability in gleaning genetic information from one species and correlating that information to another species

באפרורכות - MU עומטניספעו ו -

15

20

nowhere in this art does this problem evidence more annoying exacerbation than in the genetic sequences that encode nucleic acids and proteins. Thus, for consistency and because of the highly unpredictable nature of this art, the following invention is limited, in terms of mammals, to human GPCRs – applicability of this invention to other mammalian species, while a potential possibility, is considered beyond mere rote application.

In general, when attempting to apply common "rules" from one related protein sequence to another or from one species to another, the art has typically resorted to sequence alignment, i.e., sequences are linearized and attempts are then made to find regions of commonality between two or more sequences. While useful, this approach does not always prove to result in meaningful information. In the case of GPCRs, while the general structural motif is identical for all GPCRs. the variations in lengths of the TMs, ECs and ICs make such alignment approaches from one GPCR to another difficult at best. Thus, while it may be desirable to apply a consistent approach to, e.g., constitutive activation from one GPCR to another, because of the great diversity in sequence length, fidelity, etc from one GPCR to the next, a generally applicable, and readily successful mutational alignment approach is in essence not possible. In an analogy, such an approach is akin to having a traveler start a journey at point A by giving the traveler dozens of different maps to point B, without any scale or distance markers on any of the maps, and then asking the traveler to find the shortest and most efficient route to destination B only by using the maps. In such a situation, the task can be readily simplified by having (a) a common "placemarker" on each map, and (b) the ability to measure the distance from the place-marker to destination B - this, then, will allow the traveler to select the most efficient from starting-point A to destination B.

In essence, a feature of the invention is to provide such coordinates within human GPCRs

that readily allows for creation of a constitutively active form of the human GPCRs.

As those in the art appreciate, the transmembrane region of a cell is highly hydrophobic; thus, using standard hydrophobicity plotting techniques, those in the art are readily able to determine the TM regions of a GPCR, and specifically TM6 (this same approach is also applicable to determining the EC and IC regions of the GPCR). It has been discovered that within the TM6 region of human GPCRs, a common proline residue (generally near the middle of TM6), acts as a constitutive activation "marker." By counting 15 amino acids from the proline marker, the 16th amino acid (which is located in the IC3 loop), when mutated from its endogenous form to a non-endogenous form, leads to constitutive activation of the receptor. For convenience, we refer to this as the "Human GPCR Proline Marker" Algorithm. Although the non-endogenous amino acid at this position can be any of the amino acids, most preferably, the non-endogenous amino acid is lysine. While not wishing to be bound by any theory, we believe that this position itself is unique and that the mutation at this location impacts the receptor to allow for constitutive activation.

We note that, for example, when the endogenous amino acid at the 16th position is already lysine (as is the case with GPR4 and GPR32), then in order for X to be a non-endogenous amino acid, it must be other than lysine; thus, in those situations where the endogenous GPCR has an endogenous lysine residue at the 16th position, the non-endogenous version of that GPCR preferably incorporates an amino acid other than lysine, preferably alanine, histidine and arginine, at this position. Of further note, it has been determined that GPR4 appears to be linked to Gs and active in its endogenous form (data not shown).

Because there are only 20 naturally occurring amino acids (although the use of non-naturally occurring amino acids is also viable), selection of a particular non-endogenous amino

HINRIDOCID < WO 00551549 I 1 2

15

acid for substitution at this 16th position is viable and allows for efficient selection of a non-endogenous amino acid that fits the needs of the investigator. However, as noted, the more preferred non-endogenous amino acids at the 16th position are lysine, hisitidine, arginine and alanine, with lysine being most preferred. Those of ordinary skill in the art are credited with the ability to readily determine proficient methods for changing the sequence of a codon to achieve a desired mutation.

It has also been discovered that occasionally, but not always, the proline residue marker will be preceded in TM6 by W2 (i.e., W2P¹AA₁₅X) where W is tryptophan and 2 is any amino acid residue.

Our discovery, amongst other things, negates the need for unpredictable and complicated, sequence alignment approaches commonly used by the art. Indeed, the strength of our discovery, while an algorithm in nature, is that it can be applied in a facile manner to human GPCRs, with dexterous simplicity by those in the art, to achieve a unique and highly useful end-product, *i.e.*, a constitutively activated version of a human GPCR. Because many years and significant amounts of money will be required to determine the endogenous ligands for the human GPCRs that the Human Genome project is uncovering, the disclosed invention not only reduces the time necessary to positively exploit this sequence information, but at significant cost-savings. This approach truly validates the importance of the Human Genome Project because it allows for the utilization of genetic information to not only understand the role of the GPCRs in, *e.g.*, diseases, but also provides the opportunity to improve the human condition.

D. Screening of Candidate Compounds

10

20

1. Generic GPCR screening assay techniques

When a G protein receptor becomes constitutively active, it couples to a G protein (e.g.,

Gq, Gs, Gi, Go) and stimulates release and subsequent binding of GTP to the G protein. The G protein then acts as a GTPase and slowly hydrolyzes the GTP to GDP, whereby the receptor, under normal conditions, becomes deactivated. However, constitutively activated receptors, including the non-endogenous, human constitutively active GPCRs of the present invention, continue to exchange GDP for GTP. A non-hydrolyzable analog of GTP, [35S]GTPγS, can be used to monitor enhanced binding to G proteins present on membranes which express constitutively activated receptors. It is reported that [35S]GTPγS can be used to monitor G protein coupling to membranes in the absence and presence of ligand. An example of this monitoring, among other examples well-known and available to those in the art, was reported by Traynor and Nahorski in 1995. The preferred use of this assay system is for initial screening of candidate compounds because the system is generically applicable to all G protein-coupled receptors regardless of the particular G protein that interacts with the intracellular domain of the receptor.

B 2. Specific GPCR screening assay techniques

C Once candidate compounds are identified using the "generic" G protein-coupled receptor assay (i.e., an assay to select compounds that are agonists, partial agonists, or inverse agonists), further screening to confirm that the compounds have interacted at the receptor site is preferred. For example, a compound identified by the "generic" assay may not bind to the receptor, but may instead merely "uncouple" the G protein from the intracellular domain.

a. Gs and Gi.

Gs stimulates the enzyme adenylyl cyclase. Gi (and Go), on the other hand, inhibit this enzyme. Adenylyl cyclase catalyzes the conversion of ATP to cAMP; thus,

ENCOUNT - MU UNDSTRUTT -

20

15

constitutively activated GPCRs that couple the Gs protein are associated with increased cellular levels of cAMP. On the other hand, constitutively activated GPCRs that couple the Gi (or Go) protein are associated with decreased cellular levels of cAMP. See, generally, "Indirect Mechanisms of Synaptic Transmission," Chpt. 8, From Neuron To Brain (3rd Ed.) Nichols, J.G. et al eds. Sinauer Associates, Inc. (1992). Thus, assays that detect cAMP can be utilized to determine if a candidate compound is, e.g., an inverse agonist to the receptor (i.e., such a compound would decrease the levels of cAMP). A variety of approaches known in the art for measuring cAMP can be utilized; a most preferred approach relies upon the use of anti-cAMP antibodies in an ELISA-based format. Another type of assay that can be utilized is a whole cell second messenger reporter system assay. Promoters on genes drive the expression of the proteins that a particular gene encodes. Cyclic AMP drives gene expression by promoting the binding of a cAMP-responsive DNA binding protein or transcription factor (CREB) which then binds to the promoter at specific sites called cAMP response elements and drives the expression of the gene. Reporter systems can be constructed which have a promoter containing multiple cAMP response elements before the reporter gene, e.g., β-galactosidase or luciferase. Thus, a constitutively activated Gs-linked receptor causes the accumulation of cAMP that then activates the gene and expression of the reporter protein. The reporter protein such as βgalactosidase or luciferase can then be detected using standard biochemical assays (Chen et al. 1995). With respect to GPCRs that link to Gi (or Go), and thus decrease levels of cAMP, an approach to the screening of, e.g., inverse agonists, based upon utilization of receptors that link to Gs (and thus increase levels of cAMP) is disclosed in the Example section with respect to GPR17 and GPR30.

10

b. Go and Gq.

Gq and Go are associated with activation of the enzyme phospholipase C, which in turn hydrolyzes the phospholipid PIP₂, releasing two intracellular messengers: diacycloglycerol (DAG) and inistol 1,4,5-triphoisphate (IP₃). Increased accumulation of IP₃ is associated with activation of Gq- and Go-associated receptors. See, generally, "Indirect Mechanisms of Synaptic Transmission," Chpt. 8, From Neuron To Brain (3rd Ed.) Nichols, J.G. et al eds. Sinauer Associates, Inc. (1992). Assays that detect IP3 accumulation can be utilized to determine if a candidate compound is, e.g., an inverse agonist to a Gq- or Goassociated receptor (i.e., such a compound would decrease the levels of IP₃). Gq-associated receptors can also been examined using an AP1 reporter assay in that Gq-dependent phospholipase C causes activation of genes containing AP1 elements; thus, activated Gqassociated receptors will evidence an increase in the expression of such genes, whereby inverse agonists thereto will evidence a decrease in such expression, and agonists will evidence an increase in such expression. Commercially available assays for such detection 15 are available.

E. **Medicinal Chemistry**

10

Generally, but not always, direct identification of candidate compounds is preferably conducted in conjunction with compounds generated via combinatorial chemistry techniques, whereby thousands of compounds are randomly prepared for such analysis. Generally, the results of such screening will be compounds having unique core structures; thereafter, these compounds are preferably subjected to additional chemical modification around a preferred core structure(s) to further enhance the medicinal properties thereof. Such techniques are

- 27 -

known to those in the art and will not be addressed in detail in this patent document.

F. Pharmaceutical Compositions

Candidate compounds selected for further development can be formulated into pharmaceutical compositions using techniques well known to those in the art. Suitable pharmaceutically-acceptable carriers are available to those in the art; for example, *see* Remington's Pharmaceutical Sciences, 16th Edition, 1980, Mack Publishing Co., (Oslo et al., eds.)

G. Other Utility

15

20

Although a preferred use of the non-endogenous versions of the disclosed human GPCRs is for the direct identification of candidate compounds as inverse agonists, agonists or partial agonists (preferably for use as pharmaceutical agents), these receptors can also be utilized in research settings. For example, in vitro and in vivo systems incorporating these receptors can be utilized to further elucidate and understand the roles of the receptors in the human condition, both normal and diseased, as well understanding the role of constitutive activation as it applies to understanding the signaling cascade. A value in these non-endogenous receptors is that their utility as a research tool is enhanced in that, because of their unique features, the disclosed receptors can be used to understand the role of a particular receptor in the human body before the endogenous ligand therefor is identified. Other uses of the disclosed receptors will become apparent to those in the art based upon, *inter alia*, a review of this patent document.

EXAMPLES

The following examples are presented for purposes of elucidation, and not limitation, of the present invention. Following the teaching of this patent document that a mutational cassette may be utilized in the IC3 loop of human GPCRs based upon a position relative to a proline residue in TM6 to constitutively activate the receptor, and while specific nucleic acid

and amino acid sequences are disclosed herein, those of ordinary skill in the art are credited with the ability to make minor modifications to these sequences while achieving the same or substantially similar results reported below. Particular approaches to sequence mutations are within the purview of the artisan based upon the particular needs of the artisan.

5 Example 1

Preparation of Endogenous Human GPCRs

A variety of GPCRs were utilized in the Examples to follow. Some endogenous human GPCRs were graciously provided in expression vectors (as acknowledged below) and other endogenous human GPCRs were synthesized *de novo* using publicly-available sequence information.

1. GPR1 (GenBank Accession Number: U13666)

The human cDNA sequence for GPR1 was provided in pRcCMV by Brian O'Dowd (University of Toronto). GPR1 cDNA (1.4kB fragment) was excised from the pRcCMV vector as a NdeI-XbaI fragment and was subcloned into the NdeI-XbaI site of pCMV vector (see Figure 3). Nucleic acid (SEQ.ID.NO.: 1) and amino acid (SEQ.ID.NO.: 2) sequences for human GPR1 were thereafter determined and verified.

2. GPR4 (GenBank Accession Numbers: L36148, U35399, U21051)

The human cDNA sequence for GPR4 was provided in pRcCMV by Brian O'Dowd (University of Toronto). GPR1 cDNA (1.4kB fragment) was excised from the pRcCMV vector as an ApaI(blunted)-XbaI fragment and was subcloned (with most of the 5' untranslated region removed) into HindIII(blunted)-XbaI site of pCMV vector. Nucleic acid (SEQ.ID.NO.: 3) and amino acid (SEQ.ID.NO.: 4) sequences for human GPR4 were thereafter determined and verified.

3. GPR5 (GenBank Accession Number: L36149)

The cDNA for human GPR5 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 64°C for 1 min; and 72°C for 1.5 min. The 5' PCR primer contained an EcoRI site with the sequence: 5'-TATGAATTCAGATGCTCTAAACGTCCCTGC-3' (SEQ.ID.NO.: 5) and the 3' primer contained BamHI site with the sequence: 5'-TCCGGATCCACCTGCACCTGCGCCTGCACC-3' (SEQ.ID.NO.: 6).

The 1.1 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of PCMV expression vector. Nucleic acid (SEQ.ID.NO.: 7) and amino acid (SEQ.ID.NO.:

4. GPR7 (GenBank Accession Number: U22491)

8) sequences for human GPR5 were thereafter determined and verified.

The cDNA for human GPR7 was generated and cloned into pCMV expression vector as follows: PCR condition- PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1 min; and 72°C for 1 min and 20 sec. The 5' PCR primer contained a HindIII site with the sequence:

- 5'-GCAAGCTTGGGGGACGCCAGGTCGCCGGCT-3' (SEQ.ID.NO.: 9)
- and the 3' primer contained a BamHI site with the sequence:
 - 5'-GCGGATCCGGACGCTGGGGGAGTCAGGCTGC-3' (SEQ.ID.NO.: 10).
 - The 1.1 kb PCR fragment was digested with HindIII and BamHI and cloned into HindIII-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 11) and amino acid (SEQ.ID.NO.:

12) sequences for human GPR7 were thereafter determined and verified.

5. GPR8 (GenBank Accession Number: U22492)

The cDNA for human GPR8 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1 min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the sequence:

- 5'-CGGAATTCGTCAACGGTCCCAGCTACAATG-3' (SEQ.ID.NO.: 13).
- and the 3' primer contained a BamHI site with the sequence:
 - 5'-ATGGATCCCAGGCCCTTCAGCACCGCAATAT-3'(SEQ.ID.NO.: 14).

The 1.1 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of PCMV expression vector. All 4 cDNA clones sequenced contained a possible polymorphism involving a change of amino acid 206 from Arg to Gln. Aside from this difference, nucleic acid (SEQ.ID.NO.: 15) and amino acid (SEQ.ID.NO.: 16) sequences for human GPR8 were thereafter determined and verified.

6. GPR9 (GenBank Accession Number: X95876)

The cDNA for human GPR9 was generated and cloned into pCMV expression vector as follows: PCR was performed using a clone (provided by Brian O'Dowd) as template and pfu polymerase (Stratagene) with the buffer system provided by the manufacturer supplemented with 10% DMSO, 0.25 µM of each primer, and 0.5 mM of each of the 4 nucleotides. The cycle condition was 25 cycles of: 94°C for 1 min; 56°C for 1 min; and 72 °C for 2.5 min. The 5' PCR primer contained an EcoRI site with the sequence:

- 31 -

5'-ACGAATTCAGCCATGGTCCTTGAGGTGAGTGACCACCAAGTGCTAAAT-3' (SEQ.ID.NO.: 17)

and the 3' primer contained a BamHI site with the sequence:

5'-GAGGATCCTGGAATGCGGGGAAGTCAG-3' (SEQ.ID.NO.: 18).

The 1.2 kb PCR fragment was digested with EcoRI and cloned into EcoRI-SmaI site of PCMV expression vector. Nucleic acid (SEQ.ID.NO.: 19) and amino acid (SEQ.ID.NO.: 20) sequences for human GPR9 were thereafter determined and verified.

7. GPR9-6 (GenBank Accession Number: U45982)

The cDNA for human GPR9-6 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1 min; and 72 °C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence: 5'-TTAAGCTTGACCTAATGCCATCTTGTGTCC-3' (SEQ.ID.NO.: 21)

and the 3' primer contained a BamHI site with the sequence:

20

5'-TTGGATCCAAAAGAACCATGCACCTCAGAG-3' (SEQ.ID.NO.: 22).

The 1.2 kb PCR fragment was digested with BamHI and cloned into EcoRV-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 23) and amino acid (SEQ.ID.NO.: 24) sequences for human GPR9-6 were thereafter determined and verified.

8. GPR10 (GenBank Accession Number: U32672)

The human cDNA sequence for GPR10 was provided in pRcCMV by Brian O'Dowd (University of Toronto). GPR10 cDNA (1.3kB fragment) was excised from the pRcCMV vector as an EcoRI-XbaI fragment and was subcloned into EcoRI-XbaI site of pCMV

vector. Nucleic acid (SEQ.ID.NO.: 25) and amino acid (SEQ.ID.NO.: 26) sequences for human GPR10 were thereafter determined and verified.

9. GPR15 (GenBank Accession Number: U34806)

The human cDNA sequence for GPR15 was provided in pCDNA3 by Brian O'Dowd (University of Toronto). GPR15 cDNA (1.5kB fragment) was excised from the pCDNA3 vector as a HindIII-Bam fragment and was subcloned into HindIII-Bam site of pCMV vector. Nucleic acid (SEQ.ID.NO.: 27) and amino acid (SEQ.ID.NO.: 28) sequences for human GPR15 were thereafter determined and verified.

10. GPR17 (GenBank Accession Number: Z94154)

- The cDNA for human GPR17 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 56°C for 1 min and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the sequence:
 - 5'-CTAGAATTCTGACTCCAGCCAAAGCATGAAT-3' (SEQ.ID.NO.: 29) and the 3' primer contained a BamHI site with the sequence:
 - 5'-GCTGGATCCTAAACAGTCTGCGCTCGGCCT-3' (SEQ.ID.NO.: 30).
- The 1.1 kb PCR fragment was digested with EcoRI and BamHl and cloned into EcoRI-BamHl site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 31) and amino acid (SEQ.ID.NO.: 32) sequences for human GPR17 were thereafter determined and verified.

11. GPR18 (GenBank Accession Number: L42324)

The cDNA for human GPR18 was generated and cloned into pCMV expression

vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 54°C for 1 min; and 72 °C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence:

5'-ATAAGATGATCACCCTGAACAATCAAGAT -3' (SEQ.ID.NO.: 33)
and the 3' primer contained an EcoRI site with the sequence:

5'-TCCGAATTCATAACATTTCACTGTTTATATTGC-3' (SEQ.ID.NO.: 34).

The 1.0 kb PCR fragment was digested with EcoRI and cloned into blunt-EcoRI site of pCMV expression vector. All 8 cDNA clones sequenced contained 4 possible polymorphisms involving changes of amino acid 12 from Thr to Pro, amino acid 86 from Ala to Glu, amino acid 97 from Ile to Leu and amino acid 310 from Leu to Met. Aside from these changes, nucleic acid (SEQ.ID.NO.: 35) and amino acid (SEQ.ID.NO.: 36) sequences for human GPR18 were thereafter determined and verified.

12. GPR20 (GenBank Accession Number: U66579)

The cDNA for human GPR20 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1 min; and 72°C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence:

20 5'-CCAAGCTTCCAGGCCTGGGGTGTGCTGG-3' (SEQ.ID.NO.: 37)

and the 3' primer contained a BamHI site with the sequence:

5'-ATGGATCCTGACCTTCGGCCCCTGGCAGA-3' (SEQ.ID.NO.: 38).

The 1.2 kb PCR fragment was digested with BamHI and cloned into EcoRV-BamHI site of

PCMV expression vector. Nucleic acid (SEQ.ID.NO.: 39) and amino acid (SEQ.ID.NO.: 40) sequences for human GPR20 were thereafter determined and verified.

13. GPR21 (GenBank Accession Number: U66580)

The cDNA for human GPR21 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1 min; and 72 °C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence: 5'-GAGAATTCACTCCTGAGCTCAAGATGAACT-3' (SEQ.ID.NO.: 41)

and the 3' primer contained a BamHI site with the sequence:

5'-CGGGATCCCCGTAACTGAGCCACTTCAGAT-3' (SEQ.ID.NO.: 42).

The 1.1 kb PCR fragment was digested with BamHI and cloned into EcoRV-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 43) and amino acid (SEQ.ID.NO.: 44) sequences for human GPR21 were thereafter determined and verified.

14. GPR22 (GenBank Accession Number: U66581)

The cDNA for human GPR22 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 50°C for 1 min; and 72 °C for 1.5 min. The 5' PCR primer was kinased with the sequence:

5'-TCCCCGGGAAAAAAACCAACTGCTCCAAA-3' (SEQ.ID.NO.: 45) and the 3' primer contained a BamHI site with the sequence:

5'-TAGGATCCATTTGAATGTGGATTTGGTGAAA-3' (SEQ.ID.NO.: 46).

Tradeserving SMC Woodeday I

The 1.38 kb PCR fragment was digested with BamHI and cloned into EcoRV-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 47) and amino acid (SEQ.ID.NO.: 48) sequences for human GPR22 were thereafter determined and verified.

15. GPR24 (GenBank Accession Number: U71092)

The cDNA for human GPR24 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 56°C for 1 min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contains a HindIII site with the sequence:

5'-GTGAAGCTTGCCTCTGGTGCCTGCAGGAGG-3' (SEQ.ID.NO.: 49) and the 3' primer contains an EcoRI site with the sequence:
5'-GCAGAATTCCCGGTGGCGTGTTGTGGTGCCC-3' (SEQ.ID.NO.: 50).

The 1.3 kb PCR fragment was digested with HindIII and EcoRI and cloned into HindIII-EcoRI site of pCMV expression vector. The nucleic acid (SEQ.ID.NO.: 51) and amino acid sequence (SEQ.ID.NO.: 52) for human GPR24 were thereafter determined and verified.

16. GPR30 (GenBank Accession Number: U63917)

The cDNA for human GPR30 was generated and cloned as follows: the coding sequence of GPR30 (1128bp in length) was amplified from genomic DNA using the primers:

5'-GGCGGATCCATGGATGTGACTTCCCAA-3' (SEQ.ID.NO.: 53) and
5'-GGCGGATCCCTACACGGCACTGCTGAA-3' (SEQ.ID.NO.: 54).
The amplified product was then cloned into a commercially available vector, pCR2.1 (Invitrogen),

using a "TOPO-TA Cloning Kit" (Invitrogen, #K4500-01), following manufacturer instructions. The full-length GPR30 insert was liberated by digestion with BamH1, separated from the vector by agarose gel electrophoresis, and purified using a Sephaglas Bandprep™ Kit (Pharmacia, #27-9285-01) following manufacturer instructions. The nucleic acid (SEQ.ID.NO.: 55) and amino acid sequence (SEQ.ID.NO.: 56) for human GPR30 were thereafter determined and verified.

17. GPR31 (GenBank Accession Number: U65402)

The cDNA for human GPR31 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 58°C for 1 min; and 72 °C for 2 min. The 5' PCR primer contained an EcoRI site with the sequence: 5'-AAGGAATTCACGGCCGGGTGATGCCATTCCC-3' (SEQ.ID.NO.: 57) and the 3' primer contained a BamHI site with the sequence: 5'-GGTGGATCCATAAACACGGGCGTTGAGGAC -3' (SEQ.ID.NO.: 58).

The 1.0 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 59) and amino acid (SEQ.ID.NO.: 60) sequences for human GPR31 were thereafter determined and verified.

18. GPR32 (GenBank Accession Number: AF045764)

The cDNA for human GPR32 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerasc (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 56°C for 1 min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the

WO 00/22129 PCT/US99/23938

- 37 -

sequence:

- 5'-TAAGAATTCCATAAAAATTATGGAATGG-3' (SEQ.ID.NO.:243) and the 3' primer contained a BamHI site with the sequence:
 5'-CCAGGATCCAGCTGAAGTCTTCCATCATTC-3' (SEQ.ID.NO.: 244).
- The 1.1 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 245) and amino acid (SEQ.ID.NO.: 246) sequences for human GPR32 were thereafter determined and verified.

19. GPR40 (GenBank Accession Number: AF024687)

The cDNA for human GPR40 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min, 65°C for 1 min and 72 °C for 1 min and 10 sec. The 5' PCR primer contained an EcoRI site with the sequence

5'-GCAGAATTCGGCGGCCCCATGGACCTGCCCCC-3' (SEQ.ID.NO.: 247)
and the 3' primer contained a BamHI site with the sequence
5'-GCTGGATCCCCCGAGCAGTGGCGTTACTTC-3' (SEQ.ID.NO.: 248).
The 1 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site
of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 249) and amino acid (SEQ.ID.NO.: 250)
sequences for human GPR40 were thereafter determined and verified.

20. GPR41 (GenBank Accession Number AF024688)

The cDNA for human GPR41 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase

(Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 65°C for 1 min and 72 °C for 1 min and 10 sec. The 5' PCR primer contained an HindIII site with the sequence:

- 5'-CTCAAGCTTACTCTCTCACCAGTGGCCAC-3' (SEQ.ID.NO.: 251) and the 3' primer was kinased with the sequence 5'-CCCTCCTCCCCGGAGGACCTAGC-3' (SEQ.ID.NO.: 252).
 - The 1 kb PCR fragment was digested with HindIII and cloned into HindIII-blunt site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 253) and amino acid (SEQ.ID.NO.: 254) sequences for human GPR41 were thereafter determined and verified.

21. GPR43 (GenBank Accession Number AF024690)

The cDNA for human GPR43 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1 min; and 72 °C for 1 min and 10 sec. The 5' PCR primer contains an HindIII site with the sequence:

- 5'-TTTAAGCTTCCCCTCCAGGATGCTGCCGGAC-3' (SEQ.ID.NO.: 255) and the 3' primer contained an EcoRI site with the sequence:
- 5'-GGCGAATTCTGAAGGTCCAGGGAAACTGCTA-3' (SEQ.ID.NO. 256).
 The 1 kb PCR fragment was digested with HindIII and EcoRI and cloned into HindIII-EcoRI site
 of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 257) and amino acid (SEQ.ID.NO.: 258)
 sequences for human GPR43 were thereafter determined and verified.

BNSDOCID -WO 000010641

10

20

22. APJ (GenBank Accession Number: U03642)

Human APJ cDNA (in pRcCMV vector) was provided by Brian O'Dowd (University of Toronto). The human APJ cDNA was excised from the pRcCMV vector as an EcoRI-XbaI (blunted) fragment and was subcloned into EcoRI-SmaI site of pCMV vector. Nucleic acid (SEQ.ID.NO.: 61) and amino acid (SEQ.ID.NO.: 62) sequences for human APJ were thereafter determined and verified.

23. BLR1 (GenBank Accession Number: X68149)

The cDNA for human BLR1 was generated and cloned into pCMV expression vector as follows: PCR was performed using thymus cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1 min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the sequence:

- 5'-TGAGAATTCTGGTGACTCACAGCCGGCACAG-3' (SEQ.ID.NO.: 63):
- and the 3' primer contained a BamHI site with the sequence:
 - 5'-GCCGGATCCAAGGAAAAGCAGCAATAAAAGG-3' (SEQ.ID.NO.: 64). The 1.2 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 65) and amino acid (SEQ.ID.NO.: 66) sequences for human BLR1 were thereafter determined and verified.

24. CEPR (GenBank Accession Number: U77827)

The cDNA for human CEPR was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, $0.25~\mu\text{M}$ of each primer, and

0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1 min; and 72 °C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence: 5'-CAAAGCTTGAAAGCTGCACGGTGCAGAGAC-3' (SEQ.ID.NO.:67) and the 3' primer contained a BamHI site with the sequence:

5 5'-GCGGATCCCGAGTCACACCCTGGCTGGGCC-3' (SEQ.ID.NO.: 68).

The 1.2 kb PCR fragment was digested with BamHI and cloned into EcoRV-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 69) and amino acid (SEQ.ID.NO.: 70) sequences for human CEPR were thereafter determined and verified.

25. EBI1 (GenBank Accession Number: L31581)

The cDNA for human EBI1 was generated and cloned into pCMV expression vector as follows: PCR was performed using thymus cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1 min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the sequence:

5'-ACAGAATTCCTGTGTGGTTTTACCGCCCAG-3' (SEQ.ID.NO.: 71) and the 3' primer contained a BamHI site with the sequence:

5'-CTCGGATCCAGGCAGAAGAGTCGCCTATGG-3' (SEQ.ID.NO.: 72).

The 1.2 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of PCMV expression vector. Nucleic acid (SEQ.ID.NO.: 73) and amino acid (SEQ.ID.NO.: 74) sequences for human EBI1 were thereafter determined and verified.

26. EBI2 (GenBank Accession Number: L08177)

The cDNA for human EBI2 was generated and cloned into pCMV expression

ENSCION - NO MOSTOSAT :

vector as follows: PCR was performed using cDNA clone (graciously provided by Kevin Lynch, University of Virginia Health Sciences Center; the vector utilized was not identified by the source) as template and pfu polymerase (Stratagene) with the buffer system provided by the manufacturer supplemented with 10% DMSO, 0.25 μM of each primer, and 0.5 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 60°C for 1 min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the sequence: 5'-CTGGAATTCACCTGGACCACCAATGGATA-3' (SEQ.ID.NO.: 75) and the 3' primer contained a BamHI site with the sequence

The 1.2 kb PCR fragment was digested with EcoRI and BarnHI and cloned into EcoRI-BarnHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 77) and amino acid (SEQ.ID.NO.: 78) sequences for human EBI2 were thereafter determined and verified.

5'-CTCGGATCCTGCAAAGTTTGTCATACAG TT-3' (SEQ.ID.NO.: 76).

27. ETBR-LP2 (GenBank Accession Number: D38449)

The cDNA for human ETBR-LP2 was generated and cloned into pCMV expression vector as follows: PCR was performed using brain cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1 min; and 72 °C for 1.5 min. The 5' PCR contained an EcoRI site with the sequence:

5'-CTGGAATTCTCCTGCTCATCCAGCCATGCGG -3' (SEQ.ID.NO.: 79)
and the 3' primer contained a BamHI site with the sequence:
5'-CCTGGATCCCCACCCCTACTGGGGCCTCAG -3' (SEQ.ID.NO.: 80).
The 1.5 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI

site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 81) and amino acid (SEQ.ID.NO.: 82) sequences for human ETBR-LP2 were thereafter determined and verified.

28. GHSR (GenBank Accession Number: U60179)

The cDNA for human GHSR was generated and cloned into pCMV expression vector as follows: PCR was performed using hippocampus cDNA as template and TaqPlus Precision polymerase (Stratagene) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 68°C for 1 min; and 72°C for 1 min and 10 sec. For first round PCR, the 5' PCR primer sequence was:

- 5'-ATGTGGAACGCGACGCCCAGCG-3' (SEQ.ID.NO.: 83) and the 3' primer sequence was:
 - 5'-TCATGTATTAATACTAGATTCT-3' (SEQ.ID.NO.: 84).

Two microliters of the first round PCR was used as template for the second round PCR where the 5' primer was kinased with sequence:

- 15 5'-TACCATGTGGAACGCGACGCCCAGCGAAGAGCCGGGGT-3'(SEQ.ID.NO.:85) and the 3' primer contained an EcoRI site with the sequence:
 - 5'-CGGAATTCATGTATTAATACTAGATTCTGTCCAGGCCCG-3'(SEQ.ID.NO.:86).

The 1.1 kb PCR fragment was digested with EcoRI and cloned into blunt-EcoRI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 87) and amino acid (SEQ.ID.NO.: 88) sequences

20 for human GHSR were thereafter determined and verified.

29. GPCR-CNS (GenBank Accession Number: AFO17262)

The cDNA for human GPCR-CNS was generated and cloned into pCMV expression vector as follows: PCR was performed using brain cDNA as template and rTth

polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, $0.25 \,\mu\text{M}$ of each primer, and $0.2 \,\text{mM}$ of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1 min; and 72 °C for 2 min. The 5' PCR primer contained a HindIII site with the sequence:

5 5'-GCAAGCTTGTGCCCTCACCAAGCCATGCGAGCC-3' (SEQ.ID.NO.: 89)
and the 3' primer contained an EcoRI site with the sequence:

5'-CGGAATTCAGCAATGAGTTCCGACAGAAGC-3' (SEQ.ID.NO.: 90).

The 1.9 kb PCR fragment was digested with HindIII and EcoRI and cloned into HindIII-EcoRI site of pCMV expression vector. All nine clones sequenced contained a potential polymorphism involving a S284C change. Aside from this difference, nucleic acid (SEQ.ID.NO.: 91) and amino-

acid (SEQ.ID.NO.: 92) sequences for human GPCR-CNS were thereafter determined and verified.

30. GPR-NGA (GenBank Accession Number: U55312)

The cDNA for human GPR-NGA was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 56°C for 1 min and 72°C for 1.5 min. The 5' PCR primer contained an EcoRI site with the sequence:

- 5'-CAGAATTCAGAGAAAAAAAGTGAATATGGTTTTT-3' (SEQ.ID.NO.: 93)
- and the 3' primer contained a BamHI site with the sequence:
 - 5'-TTGGATCCCTGGTGCATAACAATTGAAAGAAT-3' (SEQ.ID.NO.: 94).
 - The 1.3 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 95) and amino acid (SEQ.ID.NO.:

96) sequences for human GPR-NGA were thereafter determined and verified.

31. H9 (GenBank Accession Number: U52219)

The cDNA for human HB954 was generated and cloned into pCMV expression vector as follows: PCR was performed using pituitary cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min, 62°C for 1 min and 72 °C for 2 min. The 5' PCR primer contains a HindIII site with the sequence: 5'-GGAAAGCTTAACGATCCCCAGGAGCAACAT-3' (SEQ.ID.NO.: 97) and the 3' primer contains a BamHI site with the sequence:

10 5'-CTGGGATCCTACGAGAGCATTTTTCACACAG-3' (SEQ.ID.NO.: 98).

The 1.9 kb PCR fragment was digested with HindIII and BamHI and cloned into HindIII-BamHI site of pCMV expression vector. When compared to the published sequences, a different isoform with 12 bp in frame insertion in the cytoplasmic tail was also identified and designated "H9b." Both isoforms contain two potential polymorphisms involving changes of amino acid P320S and amino acid G448A. Isoform H9a contained another potential polymorphism of amino acid S493N, while isoform H9b contained two additional potential polymorphisms involving changes of amino acid I502T and amino acid A532T (corresponding to amino acid 528 of isoform H9a). Nucleic acid (SEQ.ID.NO.: 99) and amino acid (SEQ.ID.NO.: 100) sequences for human H9 were thereafter determined and verified (in the section below, both isoforms were mutated in accordance with the Human GPCR Proline Marker Algorithm).

32. HB954 (GenBank Accession Number: D38449)

The cDNA for human HB954 was generated and cloned into pCMV expression

שאכטטטט אטט טענטניפעי ו י

vector as follows: PCR was performed using brain cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 58°C for 1 min and 72°C for 2 min. The 5' PCR contained a HindIII site with the sequence:

- 5'-TCCAAGCTTCGCCATGGGACATAACGGGAGCT -3' (SEQ.ID.NO.: 101)
 and the 3' primer contained an EcoRI site with the sequence:
 5'-CGTGAATTCCAAGAATTTACAATCCTTGCT -3' (SEQ.ID.NO.: 102).
- The 1.6 kb PCR fragment was digested with HindIII and EcoRI and cloned into HindIII-EcoRI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 103) and amino acid (SEQ.ID.NO.: 104) sequences for human HB954 were thereafter determined and verified:

33. HG38 (GenBank Accession Number: AF062006)

The cDNA for human HG38 was generated and cloned into pCMV expression vector as follows: PCR was performed using brain cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 56°C for 1 min and 72°C for 1 min and 30 sec. Two PCR reactions were performed to separately obtain the 5' and 3' fragment. For the 5' fragment, the 5' PCR primer contained an HindIII site with the sequence: 5'-CCCAAGCTTCGGGCACCATGGACACCTCCC-3' (SEQ.ID.NO.: 259) and the 3' primer contained a BamHisite with the sequence:

20 5'-ACAGGATCCAAATGCACAGCACTGGTAAGC-3' (SEQ.ID.NO.: 260).

This 5' 1.5 kb PCR fragment was digested with HindIII and BamHI and cloned into an HindIII-BamHI site of pCMV. For the 3' fragment, the 5' PCR primer was kinased with the sequence: 5'-CTATAACTGGGTTACATGGTTTAAC-3' (SEQ.ID.NO. 261)

and the 3' primer contained an EcoRI site with the sequence:

5'-TTTGAATTCACATATTAATTAGAGACATGG-3' (SEQ.ID.NO.: 262).

The 1.4 kb 3' PCR fragment was digested with EcoRI and subcloned into a blunt-EcoRI site of pCMV vector. The 5' and 3' fragments were then ligated together through a common EcoRV site to generate the full length cDNA clone. Nucleic acid (SEQ.ID.NO.: 263) and amino acid (SEQ.ID.NO.: 264) sequences for human HG38 were thereafter determined and verified.

34. HM74 (GenBank Accession Number: D10923)

The cDNA for human HM74 was generated and cloned into pCMV expression vector as follows: PCR was performed using either genomic DNA or thymus cDNA (pooled) as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1 min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the sequence:

5'-GGAGAATTCACTAGGCGAGGCGCTCCATC-3' (SEQ.ID.NO.: 105)

and the 3' primer was kinased with the sequence:

5'-GGAGGATCCAGGAAACCTTAGGCCGAGTCC-3' (SEQ.ID.NO.:106).

The 1.3 kb PCR fragment was digested with EcoRI and cloned into EcoRI-SmaI site of pCMV expression vector. Clones sequenced revealed a potential polymorphism involving a N94K change. Aside from this difference, nucleic acid (SEQ.ID.NO.: 107) and amino acid (SEQ.ID.NO.: 108) sequences for human HM74 were thereafter determined and verified.

35. MIG (GenBank Accession Numbers: AFO44600 and AFO44601)

The cDNA for human MIG was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and TaqPlus Precision

polymerase (Stratagene) for first round PCR or pfu polymerase (Stratagene) for second round PCR with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM (TaqPlus Precision) or 0.5 mM (pfu) of each of the 4 nucleotides. When pfu was used, 10% DMSO was included in the buffer. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1 min; and 72 °C for: (a) 1 min for first round PCR; and (b) 2 min for second round PCR. Because there is an intron in the coding region, two sets of primers were separately used to generate overlapping 5' and 3' fragments. The 5' fragment PCR primers were:

5'-ACCATGGCTTGCAATGGCAGTGCGGCCAGGGGGCACT-3' (external sense) (SEQ.ID.NO.: 109)

10 and

5'-CGACCAGGACAAACAGCATCTTGGTCACTTGTCTCCGGC-3'(internal antisense) (SEQ.ID.NO.: 110).

The 3' fragment PCR primers were:

5'-GACCAAGATGCTGTTTGTCCTGGTCGTGGTGTTTTGGCAT-3' (internal sense)

5 (SEQ.ID.NO.: 111) and

5'-CGGAATTCAGGATCGGTCTCTTGCTGCGCCT-3' (external antisense with an EcoRI site) (SEQ.ID.NO.: 112).

The 5' and 3' fragments were ligated together by using the first round PCR as template and the kinased external sense primer and external antisense primer to perform second round PCR. The 1.2 kb PCR fragment was digested with EcoRI and cloned into the blunt-EcoRI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 113) and amino acid (SEQ.ID.NO.: 114) sequences for human MIG were thereafter determined and verified.

36. OGR1 (GenBank Accession Number: U48405)

The cDNA for human OGR1 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C

- 5 for 1min; and 72 °C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence:
 - 5'-GGAAGCTTCAGGCCCAAAGATGGGGAACAT-3' (SEQ.ID.NO.: 115):

and the 3' primer contained a BamHI site with the sequence:

5'-GTGGATCCACCCGCGGAGGACCCAGGCTAG -3' (SEQ.ID.NO.: 116).

The 1.1 kb PCR fragment was digested with BamHI and cloned into the EcoRV-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 117) and amino acid (SEQ.ID.NO.: 118) sequences for human OGR1 were thereafter determined and verified.

37. Serotonin 5HT_{2A}

The cDNA encoding endogenous human 5HT_{2A} receptor was obtained by RT-PCR using human brain poly-A⁺ RNA; a 5' primer from the 5' untranslated region with an Xho I restriction site:

5'-GACCTCGAGTCCTTCTACACCTCATC-3' (SEQ.ID.NO: 119)

and a 3' primer from the 3' untranslated region containing an Xba I site:

5'-TGCTCTAGATTCCAGATAGGTGAAAACTTG-3' (SEQ.ID.NO: 120)

PCR was performed using either TaqPlusTM precision polymerase (Stratagene) or rTthTM
polymerase (Perkin Elmer) with the buffer system provided by the manufacturers, 0.25 μM of each
primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for
1 min; 57°C for 1min; and 72°C for 2 min. The 1.5 kb PCR fragment was digested with Xba 1
and subcloned into Eco RV-Xba I site of pBluescript. The resulting cDNA clones were fully

15

5

sequenced and found to encode two amino acid changes from the published sequences. The first one was a T25N mutation in the N-terminal extracellular domain; the second is an H452Y mutation. Because cDNA clones derived from two independent PCR reactions using Taq polymerase from two different commercial sources (TaqPlusTM from Stratagene and rTthTM Perkin Elmer) contained the same two mutations, these mutations are likely to represent sequence polymorphisms rather than PCR errors. With these exceptions, the nucleic acid (SEQ.ID.NO.: 121) and amino acid (SEQ.ID.NO.: 122) sequences for human 5HT_{2A} were thereafter determined and verified.

38. Serotonin 5HT_{2C}

- The cDNA encoding endogenous human 5HT_{2C} receptor was obtained from human brain poly-A⁺ RNA by RT-PCR. The 5' and 3' primers were derived from the 5' and 3' untranslated regions and contained the following sequences:
 - 5'-GACCTCGAGGTTGCTTAAGACTGAAGC-3' (SEQ.ID.NO.: 123)
 - 5'-ATTTCTAGACATATGTAGCTTGTACCG-3' (SEQ.ID.NO.: 124)
- 5 Nucleic acid (SEQ.ID.NO.: 125) and amino acid (SEQ.ID.NO.: 126) sequences for human 5HT_{2C} were thereafter determined and verified.

39. V28 (GenBank Accession Number: U20350)

The cDNA for human V28 was generated and cloned into pCMV expression vector as follows: PCR was performed using brain cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1min; and 72°C for 1 min and 20 sec. The 5' PCR primer contained a HindIII site with the sequence: 5'-GGTAAGCTTGGCAGTCCACGCCAGGCCTTC-3' (SEQ.ID.NO.: 127)

and the 3' primer contained an EcoRI site with the sequence:

5'-TCCGAATTCTCTGTAGACACAAGGCTTTGG-3' (SEQ.ID.NO.: 128)

The 1.1 kb PCR fragment was digested with HindIII and EcoRI and cloned into HindIII-EcoRI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 129) and amino acid (SEQ.ID.NO.:

5 130) sequences for human V28 were thereafter determined and verified.

Example 2
PREPARATION OF NON-ENDOGENOUS HUMAN GPCRS

1. Site-Directed Mutagenesis

Mutagenesis based upon the Human GPCR Proline Marker approach disclosed herein was

performed on the foregoing endogenous human GPCRs using Transformer Site-Directed

Mutagenesis Kit (Clontech) according to the manufacturer instructions. For this mutagenesis

approach, a Mutation Probe and a Selection Marker Probe (unless otherwise indicated, the probe

of SEQ.ID.NO.: 132 was the same throughout) were utilized, and the sequences of these for the

specified sequences are listed below in Table B (the parenthetical number is the SEQ. ID.NO.).

For convenience, the codon mutation incorporated into the human GPCR is also noted, in standard

form:

Table B

	Receptor Identifier (Codon Mutation)	Mutation Probe Sequence (5'-3') (SEQ.ID.NO.)	Selection Marker Probe Sequence (5'-3') (SEQ.ID.NO.)
20	GPR1 (F245K)	GATCTCCAGTAGGCAT <u>AAG</u> T GGACAATTCTGG (131)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAG (132)
	GPR4 (K223A)	AGAAGGCCAAGATCGCGCGG CTGGCCCTCA (133)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
25	GPR5 (V224K)	CGGCGCCACCGCACGAAAAA GCTCATCTTC	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT

WO 00/22129

	(134)	
GPR7	GCCAAGAAGCGGGTG <u>AAG</u> TT	
(T250K)	CCTGGTGGTGGCA (135)	TGTCAGAAGT
GPR8	CAGGCGGAAGGTG <u>AAA</u> GTCC	CTCCTTCGGTCCTCCTATCG
(T259K)	TGGTCCTCGT	TGTCAGAAGT
	(136)	
GPR9	CGGCGCCTGCGGGCCAAGCG	CTCCTTCGGTCCTCCTATCG
(M254K)	GCTGGTGGTG (137)	TGTCAGAAGT
GPR9-6	CCAAGCACAAAGCC <u>AAG</u> AAA	CTCCTTCGGTCCTCCTATCG
(L241K)	GTGACCATCAC (138)	TGTCAGAAGT
GPR10	GCGCCGGCGCACC <u>AAA</u> TGCT	CTCCTTCGGTCCTCCTATCG
(F276K)	TGCTGGTGGT (139)	TGTCAGAAGT
GPR15	CAAAAAGCTGAAGAAATCT <u>A</u>	CTCCTTCGGTCCTCCTATCG
(I240K)	AGAAGATCATCTTTATTGTCG (140)	TGTCAGAAGT
GPR17	CAAGACCAAGGCA <u>AAA</u> CGCA	CTCCTTCGGTCCTCCTATCG
(V234K)	TGATCGCCAT (141)	TGTCAGAAGT
GPR18	GTCAAGGAGAAGTCC <u>AAA</u> AG	CTCCTTCGGTCCTCCTATCG
(I231K)	GATCATCATC (142)	TGTCAGAAGT
GPR20	CGCCGCGTGCGGGCCAAGCA	CTCCTTCGGTCCTCCTATCG
(M240K)	GCTCCTGCTC (143)	TGTCAGAAGT
GPR21	CCTGATAAGCGCTAT <u>AAA</u> AT	CTCCTTCGGTCCTCCTATCG
(A251K)	GGTCCTGTTTCGA (144)	TGTCAGAAGT
GPR22	GAAAGACAAAAGAGAGTC <u>A</u>	CTCCTTCGGTCCTCCTATCG
(F312K)	AGAGGATGTCTTTATTG (145)	TGTCAGAAGT
GPR24	CGGAGAAAGAGGGTG <u>AAA</u> C	CTCCTTCGGTCCTCCTATCG
(T304K)	GCACAGCCATCGCC (146)	TGTCAGAAGT
GPR30 (L258K)	alternate approach; see below	alternate approach; see belo
GPR31	AAGCTTCAGCGGGCCAAGGC	CTCCTTCGGTCCTCCTATCG
(Q221K)	ACTGGTCACC (147)	TGTCAGAAGT
GPR32	CATGCCAACCGGCCCGCGAG	ACCAGCAGCAGCCTCGCGG
(K255A)	GCTGCTGCTGGT	CCGGTTGGCATG
GPR40	(279) CGGAAGCTGCGGGCCAAATG	(280) CTCCTTCGGTCCTCCTATCGT
(A223K)	GGTGGCCGGC (265)	TGTCAGAAGT
GPR41	CAGAGGAGGGTGAAGGGGCT	CTCCTTCGGTCCTCCTATCGT
••••	GTTGGCG	TGTCAGAAGT

(A223K)	(266)	
GPR43	GGCGGCGCGAGCCAAGGGG	CTCCTTCGGTCCTCCTATCGT
(V221K)	CTGGCTGTGG (267)	TGTCAGAAGT
APJ 5 (L247K)	alternate approach; see below	alternate approach; see below
BLR1	CAGCGGCAGAAGGCA <u>AAA</u> A	CTCCTTCGGTCCTCCTATCGT
(V258K)	GGGTGGCCATC (148)	TGTCAGAAGT
CEPR (L258K)	CGGCAGAAGCGCAT GATCCTCGCG (149)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
		I
(I262K)	GAGCGCAACAAGGCC <u>AAA</u> A AGGTGATCATC (150)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
EBI2 (L243K)	GGTGTAAACAAAAAGGCT <u>AA</u> AAACACAATTATTCTTATT (151)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
ETBR-LP2 5 (N358K)	GAGAGCCAGCTCAAGAGCAC CGTGGTG (152)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GHSR (V262K)	CCACAAGCAAACC <u>AAG</u> AAAA TGCTGGCTGT	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
CDCD CNIC	(153)	
GPCR-CNS (N491K)	CTAGAGAGTCAGATGAAGTG TACAGTAGTGGCAC (155)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
Ø GPR-NGA	CGGACAAAGTGAAAACTAA	CTCCTTCGGTCCTCCTATCGT
(I275K)	AAAGATGTTCCTCATT (156)	TGTCAGAAGT
H9a and H9b (F236K)	GCTGAGGTTCGCAAT <u>AAA</u> CT AACCATGTTTGTG (157)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
HB954 (H265K)	GGGAGGCCGAGCTG <u>AAA</u> GCC ACCCTGCTC	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
	(158)	
HG38 (V765K)	GGGACTGCTCTATG <u>AAA</u> AAA CACATTGCCCTG (268)	CATCAAGTGTATCATGTGCC AAGTACGCCC (154)
HM74 (I230K)	CAAGATCAAGAGAGCC <u>AAA</u> A CCTTCATCATG (159)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
) MIG (T273K)	CCGGAGACAAGTG <u>AAG</u> AAG ATGCTGTTTGTC (160)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
OGR1 (Q227K)	GCAAGGACCAGATC <u>AAG</u> CGG CTGGTGCTCA (161)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
Serotonin 5HT _{2A} (C322K)	alternate approach; see below	alternate approach; see below
Serotonin 5HT _{2C} (S310K)	alternate approach; see below	alternate approach; see below

V28	CAAGAAAGCCAAAGCCAAG	CTCCTTCGGTCCTCCTATCGT
(1230K)	AAACTGATCCTTCTG	TGTCAGAAGT
(123012)	(162)	

The non-endogenous human GPCRs were then sequenced and the derived and verified nucleic acid and amino acid sequences are listed in the accompanying "Sequence Listing" appendix

to this patent document, as summarized in Table C below:

Table C

GPR1 (F245K) SEQ.ID.NO.: 163 SEQ.ID.NO. 10 GPR4 (K223A) SEQ.ID.NO.: 165 SEQ.ID.NO. GPR5 (V224K) SEQ.ID.NO.: 167 SEQ.ID.NO. GPR7 SEQ.ID.NO.: 169 SEQ.ID.NO. 15 (T250K) SEQ.ID.NO.: 171 SEQ.ID.NO. GPR8 (T259K) SEQ.ID.NO.: 171 SEQ.ID.NO. GPR9 (M254K) SEQ.ID.NO.: 173 SEQ.ID.NO. 20 GPR9-6 (L241K) SEQ.ID.NO.: 175 SEQ.ID.NO. GPR10 (F276K) SEQ.ID.NO.: 177 SEQ.ID.NO. GPR15 SEQ.ID.NO.: 179 SEQ.ID.NO. 25 (1240K) SEQ.ID.NO.: 181 SEQ.ID.NO.	: 166
(F245K) 10 GPR4 (K223A) GPR5 (V224K) GPR7 SEQ.ID.NO.: 167 SEQ.ID.NO. 15 (T250K) GPR8 (T259K) GPR9 (M254K) 20 GPR9-6 (L241K) GPR10 (F276K) GPR15 SEQ.ID.NO.: 179 SEQ.ID.NO. SEQ.ID.NO.: 177 SEQ.ID.NO. SEQ.ID.NO. 177 SEQ.ID.NO.	: 166
SEQ.ID.NO.: 165 SEQ.ID.NO.	
(K223A) SEQ.ID.NO.: 167 SEQ.ID.NO. (V224K) SEQ.ID.NO.: 169 SEQ.ID.NO. GPR7 SEQ.ID.NO.: 169 SEQ.ID.NO. 15 (T250K) SEQ.ID.NO.: 171 SEQ.ID.NO. GPR8 SEQ.ID.NO.: 171 SEQ.ID.NO. GPR9 SEQ.ID.NO.: 173 SEQ.ID.NO. (M254K) SEQ.ID.NO.: 175 SEQ.ID.NO. GPR10 SEQ.ID.NO.: 177 SEQ.ID.NO. GPR15 SEQ.ID.NO.: 179 SEQ.ID.NO. 25 (I240K) SEQ.ID.NO.: 181 SEQ.ID.NO.	
GPR5 (V224K) SEQ.ID.NO.: 167 SEQ.ID.NO. GPR7 SEQ.ID.NO.: 169 SEQ.ID.NO. 15 (T250K) SEQ.ID.NO.: 171 SEQ.ID.NO. GPR8 (T259K) SEQ.ID.NO.: 171 SEQ.ID.NO. GPR9 (M254K) SEQ.ID.NO.: 173 SEQ.ID.NO. 20 GPR9-6 (L241K) SEQ.ID.NO.: 175 SEQ.ID.NO. GPR10 (F276K) SEQ.ID.NO.: 177 SEQ.ID.NO. GPR15 (I240K) SEQ.ID.NO.: 179 SEQ.ID.NO. 25 (I240K) SEQ.ID.NO.: 181 SEQ.ID.NO.	: 168
(V224K) GPR7 SEQ.ID.NO.: 169 SEQ.ID.NO. 15 (T250K) SEQ.ID.NO.: 171 SEQ.ID.NO. GPR8 (T259K) SEQ.ID.NO.: 173 SEQ.ID.NO. GPR9 (M254K) SEQ.ID.NO.: 173 SEQ.ID.NO. 20 GPR9-6 (L241K) SEQ.ID.NO.: 175 SEQ.ID.NO. GPR10 (F276K) SEQ.ID.NO.: 177 SEQ.ID.NO. GPR15 (SEQ.ID.NO.: 179 SEQ.ID.NO. 25 (I240K) SEQ.ID.NO.: 181 SEQ.ID.NO.	: 168
GPR7 SEQ.ID.NO.: 169 SEQ.ID.NO. 15 (T250K) GPR8 SEQ.ID.NO.: 171 SEQ.ID.NO. GPR9 SEQ.ID.NO.: 173 SEQ.ID.NO. (M254K) SEQ.ID.NO.: 175 SEQ.ID.NO. 20 GPR9-6 (L241K) SEQ.ID.NO.: 177 SEQ.ID.NO. GPR10 (F276K) SEQ.ID.NO.: 177 SEQ.ID.NO. GPR15 SEQ.ID.NO.: 179 SEQ.ID.NO. 25 (I240K) GPR17 SEQ.ID.NO.: 181 SEQ.ID.NO.	
15 (T250K) GPR8 (T259K) GPR9 (M254K) 20 GPR9-6 (L241K) GPR10 (F276K) GPR15 SEQ.ID.NO.: 177 SEQ.ID.NO. SEQ.ID.NO.: 177 SEQ.ID.NO. SEQ.ID.NO.: 177 SEQ.ID.NO. SEQ.ID.NO.: 177 SEQ.ID.NO. SEQ.ID.NO.: 179 SEQ.ID.NO. SEQ.ID.NO.: 179 SEQ.ID.NO. SEQ.ID.NO.: 179 SEQ.ID.NO. SEQ.ID.NO.: 179 SEQ.ID.NO. SEQ.ID.NO.: 181	,
GPR8 (T259K) GPR9 SEQ.ID.NO.: 173 SEQ.ID.NO. (M254K) 20 GPR9-6 SEQ.ID.NO.: 175 SEQ.ID.NO. (L241K) GPR10 SEQ.ID.NO.: 177 SEQ.ID.NO. (F276K) GPR15 SEQ.ID.NO.: 179 SEQ.ID.NO. 25 (I240K) GPR17 SEQ.ID.NO.: 181 SEQ.ID.NO.	: 170
(T259K) SEQ.ID.NO.: 173 SEQ.ID.NO. GPR9 (M254K) SEQ.ID.NO.: 175 SEQ.ID.NO. 20 GPR9-6 (L241K) SEQ.ID.NO.: 175 SEQ.ID.NO. GPR10 (F276K) SEQ.ID.NO.: 177 SEQ.ID.NO. GPR15 SEQ.ID.NO.: 179 SEQ.ID.NO. SEQ.ID.NO. 25 (I240K) SEQ.ID.NO.: 181 SEQ.ID.NO.	
GPR9 SEQ.ID.NO.: 173 SEQ.ID.NO. 20 GPR9-6 SEQ.ID.NO.: 175 SEQ.ID.NO. 20 GPR10 SEQ.ID.NO.: 177 SEQ.ID.NO. 25 (1240K) GPR15 SEQ.ID.NO.: 179 SEQ.ID.NO. 25 (1240K) GPR17 SEQ.ID.NO.: 181 SEQ.ID.NO.	: 172
(M254K) 20 GPR9-6 (L241K) GPR10 (F276K) GPR15 SEQ.ID.NO.: 177 SEQ.ID.NO. SEQ.ID.NO. 179 SEQ.ID.NO. 25 (1240K) GPR17 SEQ.ID.NO.: 181 SEQ.ID.NO.	
20 GPR9-6 (L241K) GPR10 SEQ.ID.NO.: 177 SEQ.ID.NO. (F276K) GPR15 SEQ.ID.NO.: 179 SEQ.ID.NO. 25 (I240K) GPR17 SEQ.ID.NO.: 181 SEQ.ID.NO.	: 174
(L241K) GPR10 (F276K) GPR15 SEQ.ID.NO.: 177 SEQ.ID.NO. SEQ.ID.NO. 3 (1240K) GPR17 SEQ.ID.NO.: 181 SEQ.ID.NO.	
GPR10 SEQ.ID.NO.: 177 SEQ.ID.NO. (F276K) GPR15 SEQ.ID.NO.: 179 SEQ.ID.NO. 25 (1240K) GPR17 SEQ.ID.NO.: 181 SEQ.ID.NO.	: 176
(F276K) GPR15 SEQ.ID.NO.: 179 SEQ.ID.NO. 25 (1240K) GPR17 SEQ.ID.NO.: 181 SEQ.ID.NO.	
GPR15 SEQ.ID.NO.: 179 SEQ.ID.NO. 25 (1240K) SEQ.ID.NO.: 181 SEQ.ID.NO.	: 178
25 (1240K) SEQ.ID.NO.: 181 SEQ.ID.NO.	
GPR17 SEQ.ID.NO.: 181 SEQ.ID.NO.	: 180
1	
	: 182
(V234K)	
GPR18 SEQ.ID.NO.: 183 SEQ.ID.NO.	: 184
(1231K)	
30 GPR20 SEQ.ID.NO.: 185 SEQ.ID.NO.	: 186
(M240K)	
GPR21 SEQ.ID.NO.: 187 SEQ.ID.NO.	: 188
(A251K)	
GPR22 SEQ.ID.NO.: 189 SEQ.ID.NO.	: 190
35 (F312K)	·
GPR24 SEQ.ID.NO.: 191 SEQ.ID.NO.	
(T304K))	: 192
GPR30 SEQ.ID.NO.: 193 SEQ.ID.NO.	

	(L258K)		
	GPR31	SEQ.ID.NO.: 195	SEQ.ID.NO.: 196
	(Q221K)		522.25.210 150
	GPR32	SEQ.ID.NO.: 269	SEQ.ID.NO.: 270
	(K255A)		
	GPR40	SEQ.ID.NO.: 271	SEQ.ID.NO.: 272
	(A223K)		
	GPR41	SEQ.ID.NO.: 273	SEQ.ID.NO.: 274
. 1	(A223K) GPR43		
1,	(V221K)	SEQ.ID.NO.: 275	SEQ.ID.NO.: 276
	APJ	SEQ.ID.NO.: 197	SEC AD MO 100
	(L247K)	3EQ.ID.NO.: 197	SEQ.ID.NO.: 198
	BLR1	SEQ.ID.NO.: 199	SEQ.ID.NO.: 200
1:	(V258K)	524.25.1(51)	5EQ.ID.110 200
	CEPR	SEQ.ID.NO.: 201	SEQ.ID.NO.: 202
	(L258K)		
	EBI1	SEQ.ID.NO.: 203	SEQ.ID.NO.: 204
20	(1262K) EBI2	670 77	
21	(L243K)	SEQ.ID.NO.: 205	SEQ.ID.NO.: 206
	ETBR-LP2	SEQ.ID.NO.: 207	SEC ID NO. 200
	(N358K)	SEQ.ID.NO.: 207	SEQ.ID.NO.: 208
	GHSR	SEQ.ID.NO.: 209	SEQ.ID.NO.: 210
25	(V262K)	520.23.7(0 20)	
	GPCR-CNS	SEQ.ID.NO.: 211	SEQ.ID.NO.: 212
	(N491K)		
	GPR-NGA	SEQ.ID.NO.: 213	SEQ.ID.NO.: 214
20	(I275K)) H9a	GEO ED VIO	
3	(F236K)	SEQ.ID.NO.: 215	SEQ.ID.NO.: 216
	H9b	SEQ.ID.NO.: 217	SEQ.ID.NO.: 218
	(F236K)	SEQ.ID.INO 217	SEQ.ID.NO.: 218
- 1	HB954	SEQ.ID.NO.: 219	SEQ.ID.NO.: 220
35	(H265K)		
Ī	HG38	SEQ.ID.NO.: 277	SEQ.ID.NO.: 278
	(V765K)		
	HM74	SEQ.ID.NO.: 221	SEQ.ID.NO.: 222
4	(I230K) MIG	OFO FD NO. 222	
70	(T273K)	SEQ.ID.NO.: 223	SEQ.ID.NO.: 224
}	OGR1	SEQ.ID.NO.: 225	CEO ID NO. 226
	(Q227K)	SEQ.ID.110 223	SEQ.ID.NO.: 226
ı	Serotonin 5HT _{2A}	SEQ.ID.NO.: 227	SEQ.ID.NO.: 228
45	(C322K)		224.11.110 220
}	Serotonin 5HT _{2C}	SEQ.ID.NO.: 229	SEQ.ID.NO.: 230
	(S310K)	JEQ.10 227	25C/TD/14O/1 520
t	V28	SEQ.ID.NO.: 231	SEQ.ID.NO.: 232
	(I230K)	524.20.10.1251	5LQ.ID.140 232
L	(25046)		

1,12

2. Alternate Mutation Approaches for Employment of the Proline Marker Algorithm: APJ; Serotonin 5HT_{2A}; Serotonin 5HT_{2C}; and GPR30

Although the above site-directed mutagenesis approach is particularly preferred, other approaches can be utilized to create such mutations; those skilled in the art are readily credited with selecting approaches to mutating a GPCR that fits within the particular needs of the artisan.

a. APJ

Preparation of the non-endogenous, human APJ receptor was accomplished by mutating L247K. Two oligonucleotides containing this mutation were synthesized:

5'- GGCTTAAGAGCATCATCGTGGTGCTGGTG-3' (SEQ.ID.NO.: 233)

5'-GTCACCACCAGCACCACGATGATGCTCTTAAGCC-3' (SEQ.ID.NO.: 234)

The two oligonucleotides were annealed and used to replace the NaeI-BstEII fragment of human, endogenous APJ to generate the non-endogenous, version of human APJ.

b. Serotonin 5HT₂₄

cDNA containing the point mutation C322K was constructed by utilizing the restriction

15 enzyme site Sph I which encompasses amino acid 322. A primer containing the C322K mutation:

5'-CAAAGAAAGTACTGGGCATCGTCTTCTTCCT-3' (SEQ.ID.NO: 235)

was used along with the primer from the 3' untranslated region of the receptor:

5'-TGCTCTAGATTCCAGATAGGTGAAAA CTTG-3' (SEQ.ID.NO.: 236)

to perform PCR (under the conditions described above). The resulting PCR fragment was then

20 used to replace the 3' end of endogenous 5HT_{2A} cDNA through the T4 polymerase blunted Sph

c. Serotonin 5HT_{2C}

I site.

The cDNA containing a S310K mutation was constructed by replacing the Sty I restriction fragment containing amino acid 310 with synthetic double stranded oligonucleotides that encode

the desired mutation. The sense strand sequence utilized had the following sequence:

5'-CTAGGGGCACCATGCAGGCTATCAACAATGAAAGAAAAGCTAAGAAAGTC-3' (SEQ. ID.NO.: 237)

and the antisense strand sequence utilized had the following sequence:

5'-CAAGGACTTTCTTAGCTTTTCATTGTTGATAGCCTGCATGGTGCCC-3' (SEQ. ID. NO.: 238)

d. GPR30

Prior to generating non-endogenous GPR30, several independent pCR2.1/GPR30 isolates were sequenced in their entirety in order to identify clones with no PCR-generated mutations. A clone having no mutations was digested with EcoR1 and the endogenous GPR30 cDNA fragment was transferred into the CMV-driven expression plasmid pCI-neo (Promega), by digesting pCI-Neo with EcoR1 and subcloning the EcoR1-liberated GPR30 fragment from pCR2.1/GPR30, to generate pCI/GPR30. Thereafter, the leucine at codon 258 was mutated to a lysine using a Quick-ChangeTM Site-Directed Mutagenesis Kit (Stratagene, #200518), according to manufacturer's instructions, and the following primers:

5'-CGGCGGCAGAAGGCGAAACGCATGATCCTCGCGGT-3' (SEQ.ID.NO.: 239) and 5'-ACCGCGAGGATCATGCGTTTCGCCTTCTGC CGCCG-3' (SEQ.ID.NO.: 240) Example 3

Receptor (Endogenous and Mutated) Expression

20

Although a variety of cells are available to the art for the expression of proteins, it is most preferred that mammalian cells be utilized. The primary reason for this is predicated upon practicalities, i.e., utilization of, e.g., yeast cells for the expression of a GPCR, while possible,

introduces into the protocol a non-mammalian cell which may not (indeed, in the case of yeast, does not) include the receptor-coupling, genetic-mechanism and secretary pathways that have evolved for mammalian systems – thus, results obtained in non-mammalian cells, while of potential use, are not as preferred as that obtained from mammalian cells. Of the mammalian cells, COS-7, 293 and 293T cells are particularly preferred, although the specific mammalian cell utilized can be predicated upon the particular needs of the artisan.

Unless otherwise noted herein, the following protocol was utilized for the expression of the endogenous and non-endogenous human GPCRs. Table D lists the mammalian cell and number utilized (per 150mm plate) for GPCR expression.

10

15

Table D

Receptor Name	Mammalian Cell
(Endogenous or Non-	(Number Utilized)
Endogenous)	1
GPR17	293 (2 x 10 ⁴)
GPR30	293 (4 x 10 ⁴)
APJ	COS-7 (5X106)
ETBR-LP2	293 (1 x 10 ⁷)
·	293T (1 x 10 ⁷)
GHSR	293 (1 x 10 ⁷)
	293T (1 x 10 ⁷)
MIG	293 (1 x 10 ⁷)
Serotonin 5HT _{2A}	293T (1 x 10 ⁷)
Serotonin 5HT _{2c}	293T (1 x 10 ⁷)

20

25

On day one, mammalian cells were plated out. On day two, two reaction tubes were prepared (the proportions to follow for each tube are per plate): tube A was prepared by mixing 20µg DNA (e.g., pCMV vector; pCMV vector with endogenous receptor cDNA, and pCMV vector with non-endogenous receptor cDNA.) in 1.2ml serum free DMEM (Irvine Scientific,

Irvine, CA); tube B was prepared by mixing 120µl lipofectamine (Gibco BRL) in 1.2ml serum free DMEM. Tubes A and B were then admixed by inversions (several times), followed by incubation at room temperature for 30-45min. The admixture is referred to as the "transfection mixture". Plated cells were washed with 1XPBS, followed by addition of 10ml serum free DMEM. 2.4ml of the transfection mixture was then added to the cells, followed by incubation for 4hrs at 37°C/5% CO₂. The transfection mixture was then removed by aspiration, followed by the addition of 25ml of DMEM/10% Fetal Bovine Serum. Cells were then incubated at 37°C/5% CO₂. After 72hr incubation, cells were then harvested and utilized for analysis.

1. Gi-Coupled Receptors: Co-Transfection with Gs-Coupled Receptors

In the case of GPR30, it has been determined that this receptor couples the G protein Gi. Gi is known to inhibit the enzyme adenylyl cyclase, which is necessary for catalyzing the conversion of ATP to cAMP. Thus, a non-endogenous, constitutively activated form of GPR30 would be expected to be associated with decreased levels of cAMP. Assay confirmation of a non-endogenous, constitutively activated form of GPR30 directly via measurement of decreasing levels of cAMP, while viable, can be preferably measured by cooperative use of a Gs-coupled receptor. For example, a receptor that is Gs-coupled will stimulate adenylyl cyclase, and thus will be associated with an increase in cAMP. The assignee of the present application has discovered that the orphan receptor GPR6 is an endogenous, constitutively activated GPCR. GPR6 couples to the Gs protein. Thus when co-transfected, one can readily verify that a putative GPR30-mutation leads to constitutive activation thereof: *i.e.*, an endogenous, constitutively activated GPR6/endogenous, non-constitutively activated GPR30 cell will evidence an elevated level of cAMP when compared with an endogenous, constitutively active GPR6/non-endogenous, constitutively activated GPR30 (the latter evidencing a comparatively lower level of cAMP).

WO 00/22129 PCT/US99/23938

Assays that detect cAMP can be utilized to determine if a candidate compound is e.g., an inverse agonist to a Gs-associated receptor (i.e., such a compound would decrease the levels of cAMP) or a Gi-associated receptor (or a Go-associated receptor) (i.e., such a candidate compound would increase the levels of cAMP). A variety of approaches known in the art for measuring cAMP can be utilized; a preferred approach relies upon the use of anti-cAMP antibodies. Another approach, 5 and most preferred, utilizes a whole cell second messenger reporter system assay. Promoters on genes drive the expression of the proteins that a particular gene encodes. Cyclic AMP drives gene expression by promoting the binding of a cAMP-responsive DNA binding protein or transcription factor (CREB) which then binds to the promoter at specific sites called cAMP response elements and drives the expression of the gene. Reporter systems can be constructed which have a promoter containing multiple cAMP response elements before the reporter gene, e.g., \(\beta\)-galactosidase or luciferase. Thus, an activated receptor such as GPR6 causes the accumulation of cAMP which then activates the gene and expression of the reporter protein. Most preferably, 293 cells are cotransfected with GPR6 (or another Gs-linked receptor) and GPR30 (or another Gi-linked receptor) plasmids, preferably in a 1:1 ratio, most preferably in a 1:4 ratio. Because GPR6 is an endogenous, constitutively active receptor that stimulates the production of cAMP, GPR6 strongly activates the reporter gene and its expression. The reporter protein such as β-galactosidase or luciferase can then be detected using standard biochemical assays (Chen et al. 1995). Cotransfection of endogenous, constitutively active GPR6 with endogenous, non-constitutively active GPR30 evidences an increase in the luciferase reporter protein. Conversely, co-transfection of endogenous, constitutively active GPR6 with non-endogenous, constitutively active GPR30 evidences a drastic decrease in expression of luciferase. Several reporter plasmids are known and available in the art for measuring a second messenger assay. It is considered well within the

skilled artisan to determine an appropriate reporter plasmid for a particular gene expression based primarily upon the particular need of the artisan. Although a variety of cells are available for expression, mammalian cells are most preferred, and of these types, 293 cells are most preferred. 293 cells were transfected with the reporter plasmid pCRE-Luc/GPR6 and non-endogenous, constitutively activated GPR30 using a Mammalian Transfection™ Kit (Stratagene, #200285) CaPO₄ precipitation protocol according to the manufacturer's instructions (see, 28 Genomics 347 (1995) for the published endogenous GPR6 sequence). The precipitate contained 400ng reporter, 80ng CMV-expression plasmid (having a 1:4 GPR6 to endogenous GPR30 or non-endogenous GPR30 ratio) and 20ng CMV-SEAP (a transfection control plasmid encoding secreted alkaline phosphatase). 50% of the precipitate was split into 3 wells of a 96-well tissue culture dish (containing 4X10⁴ cells/well); the remaining 50% was discarded. The following morning, the media was changed. 48 hr after the start of the transfection, cells were lysed and examined for luciferase activity using a Luclite™ Kit (Packard, Cat. # 6016911) and Trilux 1450 Microbeta™ liquid scintillation and luminescence counter (Wallac) as per the vendor's instructions. The data were analyzed using GraphPad Prism 2.0a (GraphPad Software Inc.).

With respect to GPR17, which has also been determined to be Gi-linked, a modification of the foregoing approach was utilized, based upon, *inter alia*, use of another Gs-linked endogenous receptor, GPR3 (*see* 23 Genomics 609 (1994) and 24 Genomics 391 (1994)). Most preferably, 293 cells are utilized. These cells were plated-out on 96 well plates at a density of 2 x 10⁴ cells per well and were transfected using Lipofectamine Reagent (BRL) the following day according to manufacturer instructions. A DNA/lipid mixture was prepared for each 6-well transfection as follows: 260ng of plasmid DNA in 100µl of DMEM were gently mixed with 2µl of lipid in 100µl of DMEM (the 260ng of plasmid DNA consisted of 200ng of a 8xCRE-Luc

5

10

15

reporter plasmid (see below), 50ng of pCMV comprising endogenous receptor or non-endogenous receptor or pCMV alone, and 10ng of a GPRS expression plasmid (GPRS in pcDNA3 (Invitrogen)). The 8XCRE-Luc reporter plasmid was prepared as follows: vector SRIF-β-gal was obtained by cloning the rat somatostatin promoter (-71/+51) at BgIV-HindIII site in the pßgal-Basic Vector (Clontech). Eight (8) copies of cAMP response element were obtained by PCR from an adenovirus template AdpCF126CCRE8 (see 7 Human Gene Therapy 1883 (1996)) and cloned into the SRIF-β-gal vector at the Kpn-BgIV site, resulting in the 8xCRE-β-gal reporter vector. The 8xCRE-Luc reporter plasmid was generated by replacing the beta-galactosidase gene in the 8xCRE-β-gal reporter vector with the luciferase gene obtained from the pGL3-basic vector (Promega) at the HindIII-BamHI site. Following 30min. incubation at room temperature, the DNA/lipid mixture was diluted with 400 µl of DMEM and 100µl of the diluted mixture was added to each well. 100 µl of DMEM with 10% FCS were added to each well after a 4hr incubation in a cell culture incubator. The next morning the transfected cells were changed with 200 µl/well of DMEM with 10% FCS. Eight (8) hours later, the wells were changed to 100 µl /well of DMEM without phenol red, after one wash with PBS. Luciferase activity were measured the next day using the LucLiteTM reporter gene assay kit (Packard) following manufacturer instructions and read on a 1450 MicroBeta[™] scintillation and luminescence counter (Wallac).

Figure 4 evidences that constitutively active GPR30 inhibits GPR6-mediated activation of CRE-Luc reporter in 293 cells. Luciferase was measured at about 4.1 relative light units in the expression vector pCMV. Endogenous GPR30 expressed luciferase at about 8.5 relative light units, whereas the non-endogenous, constitutively active GPR30 (L258K), expressed luciferase at about 3.8 and 3.1 relative light units, respectively. Co-transfection of endogenous GPR6 with endogenous GPR30, at a 1:4 ratio, drastically increased luciferase

expression to about 104.1 relative light units. Co-transfection of endogenous GPR6 with nonendogenous GPR30 (L258K), at the same ratio, drastically decreased the expression, which is evident at about 18.2 and 29.5 relative light units, respectively. Similar results were observed with respect to GPR17 with respect to co-transfection with GPR3, as set forth in

5 Figure 5.

Example 3
ASSAYS FOR DETERMINATION OF CONSTITUTIVE ACTIVITY
OF NON-ENDOGENOUS GPCRS

A. Membrane Binding Assays

10 1. [35S]GTPγS Assay

When a G protein-coupled receptor is in its active state, either as a result of ligand binding or constitutive activation, the receptor couples to a G protein and stimulates the release of GDP and subsequent binding of GTP to the G protein. The alpha subunit of the G protein-receptor complex acts as a GTPase and slowly hydrolyzes the GTP to GDP, at which point the receptor normally is deactivated. Constitutively activated receptors continue to exchange GDP for GTP. The non-hydrolyzable GTP analog, [35S]GTPγS, can be utilized to demonstrate enhanced binding of [35S]GTPγS to membranes expressing constitutively activated receptors. The advantage of using [35S]GTPγS binding to measure constitutive activation is that: (a) it is generically applicable to all G protein-coupled receptors; (b) it is proximal at the membrane surface making it less likely to pick-up molecules which affect the intracellular cascade.

The assay utilizes the ability of G protein coupled receptors to stimulate [35S]GTPγS binding to membranes expressing the relevant receptors. The assay can, therefore, be used in the direct identification method to screen candidate compounds to known, orphan and constitutively activated G protein-coupled receptors. The assay is generic and has application

מאופרוטטונו אווט מממזטניאין י

WO 00/22129 PCT/US99/23938

to drug discovery at all G protein-coupled receptors.

10

The [35S]GTPγS assay was incubated in 20 mM HEPES and between 1 and about 20mM MgCl₂ (this amount can be adjusted for optimization of results, although 20mM is preferred) pH 7.4, binding buffer with between about 0.3 and about 1.2 nM [35S]GTPγS (this amount can be adjusted for optimization of results, although 1.2 is preferred) and 12.5 to 75 μg membrane protein (e.g. COS-7 cells expressing the receptor; this amount can be adjusted for optimization, although 75μg is preferred) and 1 μM GDP (this amount can be changed for optimization) for 1 hour. Wheatgerm agglutinin beads (25 μl; Amersham) were then added and the mixture was incubated for another 30 minutes at room temperature. The tubes were then centrifuged at 1500 x g for 5 minutes at room temperature and then counted in a scintillation counter.

A less costly but equally applicable alternative has been identified which also meets the needs of large scale screening. Flash platesTM and WallacTM scintistrips may be utilized to format a high throughput [35S]GTPγS binding assay. Furthermore, using this technique, the assay can be utilized for known GPCRs to simultaneously monitor tritiated ligand binding to the receptor at the same time as monitoring the efficacy via [35S]GTPγS binding. This is possible because the Wallac beta counter can switch energy windows to look at both tritium and 35S-labeled probes. This assay may also be used to detect other types of membrane activation events resulting in receptor activation. For example, the assay may be used to monitor 32P phosphorylation of a variety of receptors (both G protein coupled and tyrosine kinase receptors). When the membranes are centrifuged to the bottom of the well, the bound [35S]GTPγS or the 32P-phosphorylated receptor will activate the scintillant which is coated of the wells. Scinti® strips (Wallac) have been used to demonstrate this principle. In addition, the assay also has utility for measuring ligand binding to receptors using radioactively labeled ligands. In a similar manner, when the radiolabeled bound

ligand is centrifuged to the bottom of the well, the scintistrip label comes into proximity with the radiolabeled ligand resulting in activation and detection.

Representative results of graph comparing Control (pCMV), Endogenous APJ and Non-Endogenous APJ, based upon the foregoing protocol, are set forth in Figure 6.

2. Ademylyl Cyclase

A Flash PlateTM Adenylyl Cyclase kit (New England Nuclear; Cat. No. SMP004A) designed for cell-based assays was modified for use with crude plasma membranes. The Flash Plate wells contain a scintillant coating which also contains a specific antibody recognizing cAMP. The cAMP generated in the wells was quantitated by a direct competition for binding of radioactive cAMP tracer to the cAMP antibody. The following serves as a brief protocol for the measurement of changes in cAMP levels in membranes that express the receptors.

Transfected cells were harvested approximately three days after transfection. Membranes were prepared by homogenization of suspended cells in buffer containing 20mM HEPES, pH 7.4 and 10mM MgCl₂. Homogenization was performed on ice using a Brinkman PolytronTM for approximately 10 seconds. The resulting homogenate was centrifuged at 49,000 X g for 15 minutes at 4°C. The resulting pellet was then resuspended in buffer containing 20mM HEPES, pH 7.4 and 0.1 mM EDTA, homogenized for 10 seconds, followed by centrifugation at 49,000 X g for 15 minutes at 4°C. The resulting pellet can be stored at -80°C until utilized. On the day of measurement, the membrane pellet was slowly thawed at room temperature, resuspended in buffer containing 20mM HEPES, pH 7.4 and 10mM MgCL₂ (these amounts can be optimized, although the values listed herein are prefereed), to yield a final protein concentration of 0.60mg/ml (the resuspended membranes were placed on ice until use).

cAMP standards and Detection Buffer (comprising 2 µCi of tracer [125] cAMP (100 µl] to

BRIGHTON MIND MODESTALL

5

15

WO 00/22129 PCT/US99/23938

11 ml Detection Buffer) were prepared and maintained in accordance with the manufacturer's instructions. Assay Buffer was prepared fresh for screening and contained 20mM HEPES, pH 7.4, 10mM MgCl₂, 20mM (Sigma), 0.1 units/ml creatine phosphokinase (Sigma), 50 µM GTP (Sigma), and 0.2 mM ATP (Sigma); Assay Buffer can be stored on ice until utilized. The assay was initiated by addition of 50ul of assay buffer followed by addition of 50ul of membrane suspension to the NEN Flash Plate. The resultant assay mixture is incubated for 60 minutes at room temperature followed by addition of 100ul of detection buffer. Plates are then incubated an additional 2-4 hours followed by counting in a Wallac MicroBeta scintillation counter. Values of cAMP/well are extrapolated from a standard cAMP curve which is contained within each assay plate. The foregoing assay was utilized with respect to analysis of MIG.

B. Reporter-Based Assays

5

20

1. CREB Reporter Assay (Gs-associated receptors)

A method to detect Gs stimulation depends on the known property of the transcription factor CREB, which is activated in a cAMP-dependent manner. A PathDetect CREB trans-Reporting System (Stratagene, Catalogue # 219010) was utilized to assay for Gs coupled activity in 293 or 293T cells. Cells were transfected with the plasmids components of this above system and the indicated expression plasmid encoding endogenous or mutant receptor using a Mammalian Transfection Kit (Stratagene, Catalogue #200285) according to the manufacurer's instructions. Briefly, 400 ng pFR-Luc (luciferase reporter plasmid containing Gal4 recognition sequences), 40 ng pFA2-CREB (Gal4-CREB fusion protein containing the Gal4 DNA-binding domain), 80 ng CMV-receptor expression plasmid (comprising the receptor) and 20 ng CMV-SEAP (secreted alkaline phosphatase expression plasmid; alkaline phosphatase activity is measured in the media of transfected cells to control for variations in

transfection efficiency between samples) were combined in a calcium phosphate precipitate as per the Kit's instructions. Half of the precipitate was equally distributed over 3 wells in a 96-well plate, kept on the cells overnight, and replaced with fresh medium the following morning. Forty-eight (48) hr after the start of the transfection, cells were treated and assayed for luciferase activity as set forth with resepct to the GPR30 system, above. This assay was used with respect to GHSR.

2. AP1 reporter assay (Gq-associated receptors)

Ae method to detect Gq stimulation depends on the known property of Gq-dependent phospholipase C to cause the activation of genes containing AP1 elements in their promoter.

A Pathdetect AP-1 cis-Reporting System (Stratagene, Catalogue # 219073) was utilized following the protocl set forth above with respect to the CREB reporter assay, except that the components of the calcium phosphate precipitate were 410 ng pAP1-Luc, 80 ng receptor expression plasmid, and 20 ng CMV-SEAP. This assay was used with respect to ETBR-LP2

C. Intracellular IP3 Accumulation Assay

On day 1, cells comprising the serotonin receptors (endogenous and mutated) were plated onto 24 well plates, usually 1x10⁵ cells/well. On day 2 cells were transfected by firstly mixing 0.25ug DNA in 50 ul serumfree DMEM/well and 2 ul lipofectamine in 50 µl serumfree DMEM/well. The solutions were gently mixed and incubated for 15-30 min at room temperature. Cells were washed with 0.5 ml PBS and 400 µl of serum free media was mixed with the transfection media and added to the cells. The cells were then incubated for 3-4 hrs at 37°C/5%CO₂ and then the transfection media was removed and replaced with 1ml/well of regular growth media. On day 3 the cells were labeled with ³H-myo-inositol. Briefly, the media was removed the cells were washed with 0.5 ml PBS. Then 0.5 ml inositol-free/serumfree media (GIBCO BRL) was added/well with 0.25 µCi of ³H-myo-inositol / well

15

WO 00/22129 PCT/US99/23938

and the cells were incubated for 16-18 hrs o/n at 37°C/5%CO2. On Day 4 the cells were washed with 0.5 ml PBS and 0.45 ml of assay medium was added containing inositolfree/serum free media 10 µM pargyline 10 mM lithium chloride or 0.4 ml of assay medium and 50 ul of 10x ketanserin (ket) to final concentration of 10µM. The cells were then incubated for 30 min at 37°C. The cells were then washed with 0.5 ml PBSand 200 ul of fresh/icecold stop solution (1M KOH; 18 mM Na-borate; 3.8 mM EDTA) was added/well. The solution was kept on ice for 5-10 min or until cells were lysed and then neutralized by 200 µl of fresh/ice cold neutralization sol. (7.5 % HCL). The lysate was then transferred into 1.5 ml eppendorf tubes and 1 ml of chloroform/methanol (1:2) was added/tube. The solution was vortexed for 15 sec and the upper phase was applied to a Biorad AG1-X8 anion exchange resin (100-200 mesh). Firstly, the resin was washed with water at 1:1.25 W/V and 0.9 ml of upper phase was loaded onto the column. The column was washed with 10 mls of 5 mM myo-inositol and 10 ml of 5 mM Na-borate/60mM Na-formate. The inositol tris phosphates were eluted into scintillation vials containing 10 ml of scintillation cocktail with 15 2 ml of 0.1 M formic acid/ 1 M ammonium formate. The columns were regenerated by washing with 10 ml of 0.1 M formic acid/3M ammonium formate and rinsed twice with dd H₂O and stored at 4°C in water.

10

Figure 7 provides an illustration of IP3 production from the human 5-HT_{2A} receptor that incorporates the C322K mutation. While these results evidence that the Proline Mutation 20 Algorithm approach constitutively activates this receptor, for purposes of using such a receptor for screening for identification of potential therapeutics, a more robust difference would be preferred. However, because the activated receptor can be utilized for understanding and elucidating the role of constitutive activation and for the identification of compounds that

can be further examined, we believe that this difference is itself useful in differentiating between the endogenous and non-endogenous versions of the human $5 \mathrm{HT}_{2A}$ receptor.

D. Result Summary

The results for the GPCRs tested are set forth in Table E where the Per-Cent Increase indicates the percentage difference in results observed for the non-endogenous GPCR as compared to the endogenous GPCR; these values are followed by parenthetical indications as to the type of assay utilized. Additionally, the assay sytem utilized is parenthetically listed (and, in cases where different Host Cells were used, both are listed). As these results indicate, a variety of assays can be utilized to determine constitutive activity of the non-endogenous versions of the human GPCRs. Those skilled in the art, based upon the foregoing and with reference to information available to the art, are creditied with theability to selelect and/ot maximize a particular assay approach that suites the particular needs of theinvestigator.

Table E

•	-
	•

20	

Receptor Identifier	Per-Cent Difference
(Codon Mutation)	
GPR17	74.5
(V234K)	(CRE-Luc)
GPR30	71.6
(L258K)	(CREB)
APJ	49.0
(L247K)	(GTP _γ S)
ETBR-LP2	48.4(AP1-Luc - 293)
(N358K)	61.1(AP1-Luc - 293T)

GHSR	58.9(CREB - 293)	
(V262K)	35.6(CREB - 293T)	

25

MIG	39 (cAMP)
(I230K)	
Serotonin 5HT _{2A}	33.2 (IP ₃)
(C322K)	
Serotonin 5HT _{2C}	39.1(IP ₃)
(S310K)	

Example 6

Tissue Distribution of Endogenous Orphan GPCRs

Using a commercially available human-tissue dot-blot format, endogenous orphan GPCRs were probed for a determination of the areas where such receptors are localized. Except as indicate below, the entire receptor cDNA (radiolabelled) was used as the probe: radiolabeled probe was generated using the complete receptor cDNA (excised from the vector) using a Prime-It II™ Random Primer Labeling Kit (Stratagene, #300385), according to manufacturer's instructions. A human RNA Master Blot™ (Clontech, #7770-1) was hybridized with the GPCR radiolabeled probe and washed under stringent conditions according manufacturer's instructions. The blot was exposed to Kodak BioMax Autoradiography film overnight at -80°C.

Representative dot-blot format results are presented in Figure 8 for GPR1 (8A), GPR30 (8B), and APJ (8C), with results being summarized for all receptors in Table F

20

15

Table F

GPCR	Tissue Distribution
	(highest levels, relative to other tissues in
	the dot-bl t)
GPR1	Placenta, Ovary, Adrenal

GPR4	Broad; highest in Heart, Lung, Adrenal
	Thyroid, Spinal Cord
GPR5	Placenta, Thymus, Fetal Thymus
	Lesser levels in spleen, fetal spleen
GPR7	Liver, Spleen, Spinal Cord, Placenta
GPR8	No expression detected
GPR9-6	Thymus, Fetal Thymus
	Lesser levels in Small Intestine
GPR18	Spleen, Lymph Node, Fetal Spleen, Testis
GPR20	Broad
GPR21	Broad; very low abundance
GPR22	Heart, Fetal Heart
	Lesser levels in Brain
GPR30	Stomach
GPR31	Broad
BLR1	Spleen
CEPR	Stomach, Liver, Thyroid, Putamen
EBI1	Pancreas
	Lesser levels in Lymphoid Tissues
EBI2	Lymphoid Tissues, Aorta, Lung, Spinal Cord
ETBR-LP2	Broad; Brain Tissue
GPCR-CNS	Brain
	Lesser levels in Testis, Placenta
GPR-NGA	Pituitary
	Lesser levels in Brain
Н9	Pituitary
HB954	Aorta, Cerebellum
	Lesser levels in most other tissues
HM74	Spleen, Leukocytes, Bone marrow, Mammary
	Glands, Lung, Trachea
MIG	Low levels in Kidney, Liver, Pancreas, Lung,
	Spleen
ORG1	Pituitary, Stomach, Placenta
V28	Brain, Spleen, Peripheral Leukocytes

Based upon the foregoing information, it is noted that human GPCRs can also be assessed for distribution in diseased tissue; comparative assessments between "normal" and diseased tissue can then be utilized to determine the potential for over-expression or under-expression of a particular receptor in a diseased state. In those circumstances where it is desirable to utilize the non-endogenous versions of the human GPCRs for the purpose of screening to directly identify

5

10

15

20

בשברבתות השת החומבושם

WO 00/22129 PCT/US99/23938

- 71 -

candidate compounds of potential therapeutic relevance, it is noted that inverse agonists are useful in the treatment of diseases and disorders where a particular human GPCR is over-expressed, whereas agonists or partial agonists are useful in the treatment of diseases and disorders where a particular human GPCR is under-expressed.

As desired, more detailed, cellular localization of the recepotrs, using techniques well-known to those in the art (e.g., in-situ hybridization) can be utilized to identify particular cells within these tissues where the receptor of interest is expressed.

It is intended that each of the patents, applications, and printed publications mentioned in this patent document be hereby incorporated by reference in their entirety.

As those skilled in the art will appreciate, numerous changes and modifications may be made to the preferred embodiments of the invention without departing from the spirit of the invention. It is intended that all such variations fall within the scope of the invention.

10

Although a variety of expression vectors are available to those in the art, for purposes of utilization for both the endogenous and non-endogenous human GPCRs, it is most preferred that the vector utilized be pCMV. This vector has been deposited with the American Type Culture Collection (ATCC) on October 13, 1998 (10801 University Blvd., Manassas, VA 20110-2209 USA) under the provisions of the Budapest Treaty for the International Recognition of the Deposit of Microorganisms for the Purpose of patent Procedure. The vector was tested by the ATCC on _______, 1998 and determined to be viable on _______, 1998. The ATCC has assigned the following deposit number to pCMV: ______.

CLAIMS

What is claimed is:

1. A constitutively active, non-endogenous version of an endogenous human orphan G protein-coupled receptor (GPCR) comprising the following amino acid residues (carboxy-terminus to amino-terminus orientation) transversing the transmembrane-6 (TM6) and intracellular loop-3 (IC3) regions of the non-endogenous GPCR:

$P^1 AA_{15} X$

wherein:

10

15

20

- (1) P¹ is an amino acid residue located within the TM6 region of the nonendogenous GPCR, where P¹ is selected from the group consisting of (i) the endogenous orphan GPCR proline residue, and (ii) a nonendogenous amino acid residue other than proline;
- of (a) the 15 endogenous amino acid residues of the endogenous orphan GPCR, (b) 15 non-endogenous amino acid residues, and (c) a combination of 15 amino acid residues, the combination comprising at least one endogenous amino acid residue of the endogenous orphan GPCR and at least one non-endogenous amino acid residue, excepting that none of the 15 endogenous amino acid residues that are positioned within the TM6 region of the GPCR is proline; and
- (2) X is a non-endogenous amino acid residue located within the IC3 region of said non-endogenous GPCR.
- 2. The non-endogenous human GPCR of claim 1 wherein P¹ is the endogenous proline

מאופחרות שוות מחמימינאי ו

residue.

- 3. The non-endogenous human GPCR of claim1 wherein P¹ is a non-endogenous amino acid residue other than a proline residue.
- 4. The non-endogenous human GPCR of claim 1 wherein AA₁₅ are the 15 endogenous amino acid residues of the endogenous GPCR.
- 5. The non-endogenous human GPCR of claim 1 wherein X is selected from the group consisting of lysine, hisitidine, arganine and alanine residues, excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X is selected from the group consisting of histidine, arginine and alanine.
- The non-endogenous human GPCR of claim 1 wherein X is a lysine residue, excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X is an amino acid other than lysine.
- 7. The non-endogenous human GPCR of claim 4 wherein X is a lysine residue, excepting that when the endogenous amino acid in position X of said endogenous human GPCR
 15 is lysine, X is an amino acid other than lysine.
 - 8. The non-endogenous, human GPCR of claim 1 wherein P¹ is a proline residue and X is a lysine residue, excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X is an amino acid other than lysine.
 - 9. A host cell comprising the non-endogenous human GPCR of claim 1.
- 20 10. The material of claim 9 wherein said host cell is of mammalian origin.
 - 11. The non-endogenous human GPCR of claim 1 in a purified and isolated form.
 - 12. A nucleic acid sequence encoding a constitutively active, non-endogenous version of an endogenous human orphan G protein-coupled receptor (GPCR) comprising the following

nucleic acid sequence region transversing the transmembrane-6 (TM6) and intracellular loop-3 (IC3) regions of the orphan GPCR:

3'-P^{codon} (AA-codon)₁₅ X_{codon}-5'

wherein:

5

(1) P^{codon} is a nucleic acid encoding region within the TM6 region of the non-endogenous GPCR, where P^{codon} encodes an amino acid selected from the group consisting of (i) the endogenous GPCR proline residue, and (ii) a non-endogenous amino acid residue other than proline;

10

(2) (AA-codon)₁₅ are 15 codons encoding 15 amino acid residues selected from the group consisting of (a) the 15 endogenous amino acid residues of the endogenous orphan GPCR, (b) 15 non-endogenous amino acid residues, and (c) a combination of 15 amino acid residues, the combination comprising at least one endogenous amino acid residue of the endogenous orphan GPCR and at least one non-endogenous amino acid residue, excepting that none of the 15 endogenous amino acid residues that are positioned within the TM6 region of the orphan GPCR is proline; and

15

(3) X_{codon} is a nucleic acid encoding region residue located within the IC3 region of said non-endogenous human GPCR, where X_{codon} encodes a non-endogenous amino acid.

20

- 13. The nucleic acid sequence of claim 12 wherein P^{codon} encodes an endogenous proline residue.
- 14. The nucleic acid sequence of claim 12 wherein P^{codon} encodes a non-endogenous

. 5

10

amino acid residue other than a proline residue.

- 15. The nucleic acid sequence of claim 12 wherein X_{codon} encodes a non-endogenous amino acid selected from the group consisting of lysine, histidine, arginine and alanine, excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X_{codon} encodes an amino acid selected from the group consisiting of histidine, arginine and alanine.
- 16. The nucleic acid sequence of claim 13 wherein X_{codon} encodes a non-endogenous lysine amino acid excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X_{codon} encodes an amino acid selected from the group consisiting of histidine, arginine and alanine.
- 17. The nucleic acid sequence of claim 12 wherein X_{codon} is selected from the group consisting of AAA, AAG, GCA, GCG, GCC and GCU.
- 18. The nucleic acid sequence of claim 12 wherein X_{codon} is selected from the group consisting of AAA and AAG.
- 15 19. The nucleic acid sequence of claim 12 wherein P^{codon} is selected from the group consisting of CCA, CCC, CCG and CCU, and X_{codon} is selected from the group consisting of AAA and AAG.
 - 20. A vector comprising the nucleic acid sequence of claim 12.
 - 21. A plasmid comprising the nucleic acid sequence of claim 12.
- 20 22. A host cell comprising the nucleic acid sequence of claim 21.
 - 23. The nucleic acid sequence of claim 12 in a purified and isolated form.
 - 24. A method for selecting for alteration an endogenous amino acid residue within the third intracellular loop of a human G protein-coupled receptor ("GPCR"), said receptor

comprising a transmembrane 6 region and an intracellular loop 3 region, which endogenous amino acid, when altered to a non-endogenous amino acid, constitutively activates said human GPCR, comprising the following steps:

- (a) identifying an endogenous proline residue within the transmembrane 6 region of a human GPCR;
- (b) identifying, by moving in a direction of the carboxy-terminus region of said GPCR towards the amino-terminus region of said GPCR, the endogenous, 16th amino acid residue from said proline residue;
- altering the endogenous residue of step (b) to a non-endogenous amino acid residue to create a non-endogenous version of an endogenous human GPCR; and
 - (d) determining whether the non-endogenous human GPCR of step (c) is constitutively active.
- 25. The method of claim 24 wherein the amino acid residue that is two residues from said proline residue in the transmembrane 6 region, in a carboxy-terminus to aminoterminus direction, is tryptophan.
 - 26. A constitutively active, non-endogenous human GPCR produced by the process of claim 24.
- A constitutively active, non-endogenous human GPCR produced by the process ofclaim 25.
 - 28. An algorithmic approach for creating a non-endogenous, constitutively active version of an endogenous human G protein coupled receptor (GPCR), said endogenous GPCR comprising a transmembrane 6 region and an intracellular loop 3 region, the

BNSDOCID ...WO DOSSESSAL

5

.5

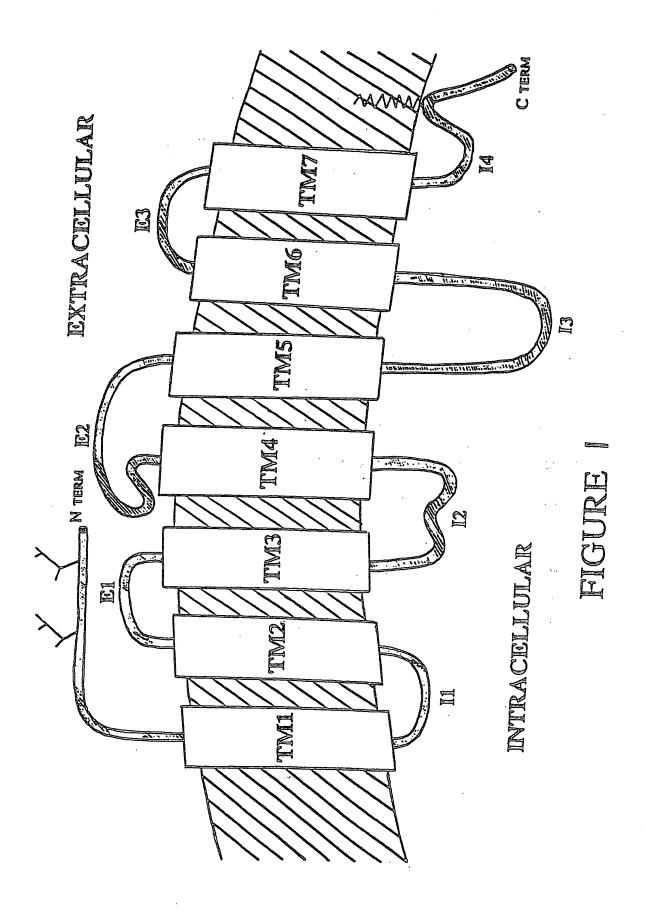
algorithmic approach comprising the steps of:

- (a) selecting an endogenous human GPCR comprising a proline residue in the transmembrane-6 region;
- (b) identifying, by counting 16 amino acid residues from the proline residue of step (a), in a carboxy-terminus to amino-terminus direction, an endogenous amino acid residue;
 - (c) altering the identified amino acid residue of step (b) to a non-endogenous amino acid residue to create a non-endogenous version of the endogenous human GPCR; and
- (d) determining if the non-endogenous version of the endogenous human GPCRof step (c) is constitutively active.
 - 29. The algorithmic approach of claim 28 wherein the amino acid residue that is two residues from said proline residue in the transmembrane 6 region, in a carboxy-terminus to amino-terminus direction, is tryptophan.
- 15 30. A constitutively active, non-endogenous human GPCR produced by the algorithmic approach of claim 28.
 - 31. A constitutively active, non-endogenous human GPCR produced by the algorithmic approach of claim 29.
- 32. A method for directly identifying a compound selected from the group consisting of inverse agonists, agonists and partial agonists to a non-endogenous, constitutively activated human G protein coupled receptor, said receptor comprising a transmembrane-6 region and an intracellular loop-3 region, comprising the steps of:
 - (a) selecting an endogenous human GPCR;

- (b) identifying a proline residue within the transmembrane-6 region of the GPCR of step (a);
- (c) identifying, in a carboxy-terminus to amino-terminus direction, the endogenous, 16th amino acid residue from the proline residue of step (b);
- 5 (d) altering the endogenous amino acid of step (c) to a non-endogenous amino acid;
 - (e) confirming that the non-endogenous GPCR of step (d) is constitutively active;
 - (f) contacting a candidate compound with the non-endogenous, constitutivelyactivated GPCR of step (e); and
- 10 (g) determining, by measurement of the compound efficacy at said contacted receptor, whether said compound is an inverse agonist, agonist or partial agonist of said receptor.
 - 33. The method of claim 32 wherein the non-endogenous amino acid of step (d) is lysine.
 - 34. A compound directly identified by the method of claim 32.
- 15 35. The method of claim 32 wherein the directly identified compound is an inverse agonist.
 - 36. The method of claim 32 wherein the directly identified compound is an agonist.--
 - 37. The method of claim 32 wherein the directly identified compound is a partial agonist.
 - 38. A composition comprising the inverse agonist of claim 35.
- 20 39. A composition comprising the agonist of claim 36.
 - 40. A composition comprising the partial agonist of claim 37.
 - 41. A method for directly identifying an inverse agonist to a non-endogenous,

constitutively activated human G protein coupled receptor ("GPCR"), said GPCR comprising a transmembrane-6 region and an intracellular loop-3 region, comprising the steps of:

- (a) selecting an endogenous human GPCR;
- (b) identifying a proline residue within the transmembrane-6 region of the GPCR of step (a);
 - (c) identifying, in a carboxy-terminus to amino-terminus direction, the endogenous, 16th amino acid residue from the proline residue of step (b);
 - (d) altering the endogenous amino acid of step (c) to a non-endogenous lysine residue;
 - (e) confirming that the non-endogenous GPCR of step (d) is constitutively active;
- (f) contacting a candidate compound with the non-endogenous, constitutivelyactivated GPCR of step (e); and
 - (g) determining, by measurement of the compound efficacy at said contacted receptor, whether said compound is an inverse agonist of said receptor.
 - 42. An inverse agonist directly identified by the method of claim 37.
- 15 43. A composition comprising an inverse agonist of claim 38.



EXTRACELLULAR

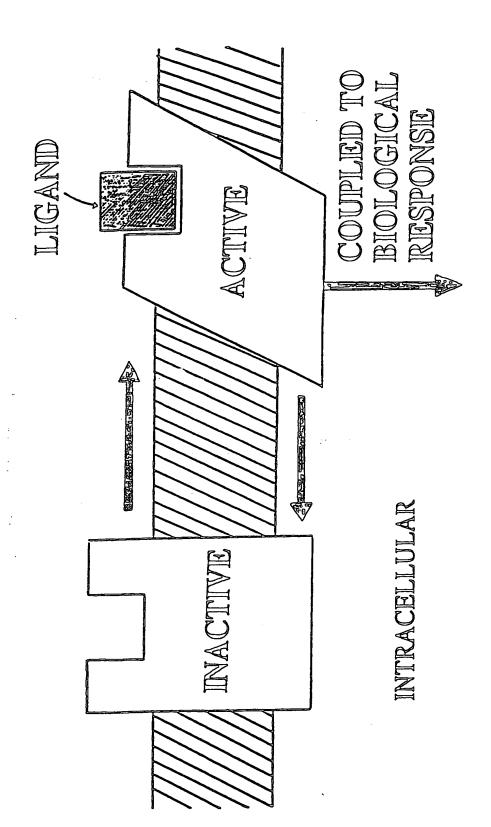
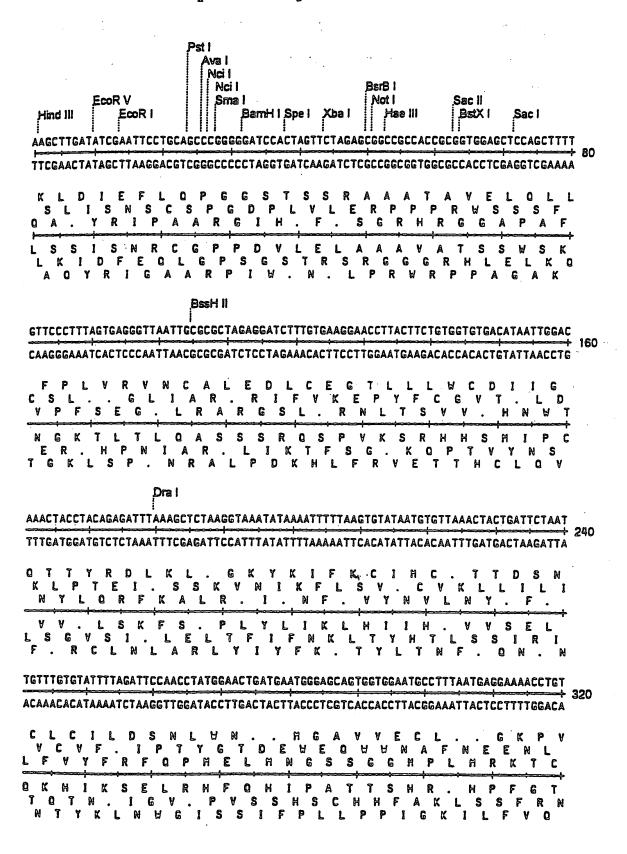


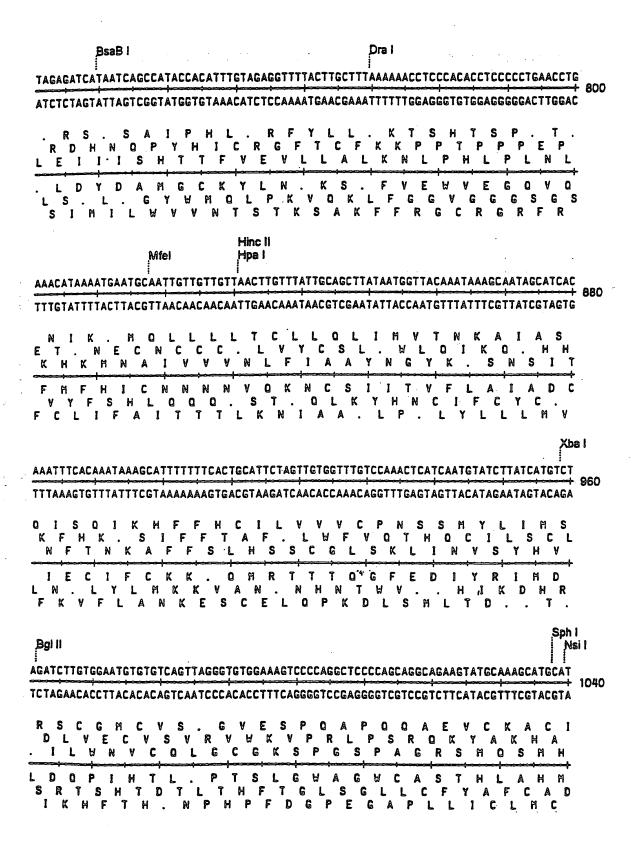
FIGURE 2

pCMV Sequence and Restriction Site



TTTGCTCAGAAGAAATGCCATCTAGTGATGATGAGGCTACTGCTGACTCTCAACATTCTACTCCTCCAAAAAAAGAAGAAGA AAACGAGTCTTCTTTACGGTAGATCACTACTACTCCGATGACGACTGAGAGTTGTAAGATGAGGAGGTTTTTTCTTCTCT PSSDDEATADSQHSTP FACKKCHL V M M R L L L T L N I L L Q K R R E KSLLFAM. HHHP. QQSEVN. EELFSSI QESSIGDLSSSAVASE. CEVGGFFFL (A. FFHWRTIILSSSVRLMRSRWFLLS SF Sty I AAGGTAGAAGACCCCAAGGACTTTCCTTCAGAATTGCTAAGTTTTTTGAGTCATGCTGTTTTAGTAATAGAACTCTTGC TTCCATCTTCTGGGGTTCCTGAAAGGAAGTCTTAACGATTCAAAAAACTCAGTACGACAAAATCATTATCTTGAGAACG CGRRPOGLSFRIAKFFESCCV...NSC KVEDPKDFPSELLSFLSHAVFSNRTLA VF. VMLCLVIELL R. KTPRTFLONC. PLLG W P S E K L I A L N K S D H Q T . Y Y F E Q F T S S G L S K G E S N S L K K L . A T N L L L V R A L Y F V G L V K R . F Q . T K Q T M S H K T I S S K S TTGCTTTGCTATTTACACCACAAAGGAAAAAGCTGCACTGCTATACAAGAAAATTATGGAAAAAATATTCTGTAACCTTTA AACGAAACGATAAATGTGGTGTTTCCTTTTTCGACGTGACGATATGTTCTTTTAATACCTTTTTATAAGACATTGGAAAT LCYLHH.KGKSCTAIQENYGK CFAIYTTKEKAALLYKKIMEKY LALLFTPORKKLHCYTRKLUKNIL. PFINOL KCWLPFLQVAICSF. V V F S F A A S S Y L F I I S F Y E T V K I AKSNVGCLFFSCQ. ٧ L FNHFFIRYGK Asel TAAGTAGGCATAACAGTTATAATCATAACATACTGTTTTTTCTTACTCCACACAGGCATAGAGTGTCTGCTATTAATAAC ATTCATCCGTATTGTCAATATTAGTATTGTATGACAAAAAAGAATGAGGTGTGTCCGTATCTCACAGACGATAATTATTG F T V F S Y S T L F F L T P H K.A.QL.S.HI ISRHNSYNHNI Н I T YCFFLLHTGIEC IIT LYAYCNYDYCYTKE. E V LLCLL L H S N K R V G C L C L T D A I L L TPH V T I I H V Y Q K K S H V P H S H R S N I V Rsa I TATGCTCAAAAATTGTGTACCTTTAGCTTTTTAATTTGTAAAGGGGTTAATAAGGAATATTTGATGTATAGTGCCTTGAC ÷ 720 ATACGAGTTTTTAACACATGGAAATCGAAAAATTAAACATTTCCCCAATTATTCCTTATAAACTACATATCACGGAACTG CL IFDV CKGVNKEYLMYSA FAKELIRNI. CIVP. YPINSTYHRS SKLKYLP HEFITYR. SKLKYLP. YPINSTYHRS A. FNHYKLKKIOLPTLLSYKIYLAKY NTFPNILFIOHITGOS

FIGURE 3B



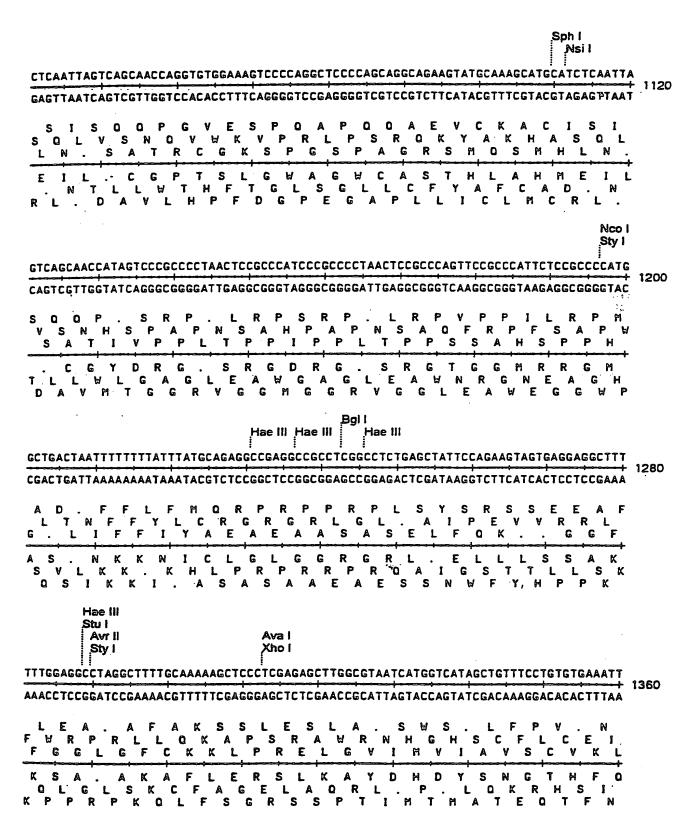
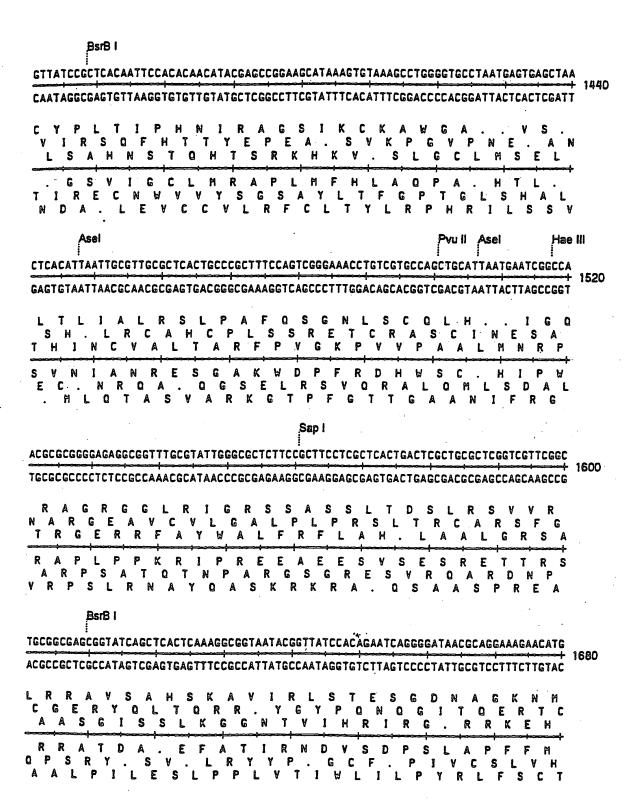


FIGURE 3D



PCT/US99/23938

Ha III Hae III Hae III TGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGA ACTCGTTTTCCGGTCGTTTTCCGGTCCTTGGCATTTTTCCGGCGCAACGACCGCAAAAAGGTATCCGAGGCGGGGGGACT SKRPGTVKRPRCWRFSI V S K R P A K G Q E P . K G R V A G V T F L G R Q Q R K E M P E A G R V Y F P R T A P T K G Y A G G G Q CGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTG GCTCGTAGTGTTTTTAGCTGCGAGTTCAGTCTCCACCGCTTTGGGCTGTCCTGATATTTCTATGGTCCGCAAAGGGGGAC EHHKNRRSSQRWRNPTGL E H H K W K K S S Q R W R N P T G L . R Y Q A F T S I T K I D A Q V R G G E T R Q D Y K D T R R F R A S Q K S T L K S E V A K P D R T I K I P G V LHRFGVPSYL LMVF1SA.TLPPSVRCS.L RADCFDVSLDSTAFGSLVIF GPTEGQ GAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTG + 1920 CTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAATGGCCTATGGACAGGCGGAAAGAGGGAAGCCCTTCGCAC LTGYLSAF EAPSCALLFRPCRLPD TCPP F S L R E A W YRIPVRLSPFGKR SCSDP RGSVPYRDAKRGEPLT SAGEHARRNRGQRKGSVQGGKERRSAH FSGRASEQESGAA.RIGTRREGKPFRP ApaL I GCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACC 2000 CGCGAAAGAGTTACGAGTGCGACATCCATAGAGTCAAGCCACATCCAGCAAGCGAGGTTCGACCCGACACACGTGCTTGG V "V R S RCRSF С SQFGVGRSL HERQLYRLETYTT'RELSPOTCS TPIETRHLD NAGLQATH V F G WAPSHARV SYTD. N P TP RES Nci I CCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCAC + 2080 GGGGCAAGTCGGGCTGGCGACGCGGAATAGGCCATTGATAGCAGAACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTG V Q P D R C A L S G N Y R L E S N P V R H D L S P VTIVLSPTR. DTT PRSARPLRLIR. LSS. VQPGKTRLIAT GSRQAKDPL RRSDLGTLCSKDGS G T V 1 T K L G V R Y S V V . R Y S D D Q T W G P L V R S LYRSI

חוווי יזויססוום

000010041

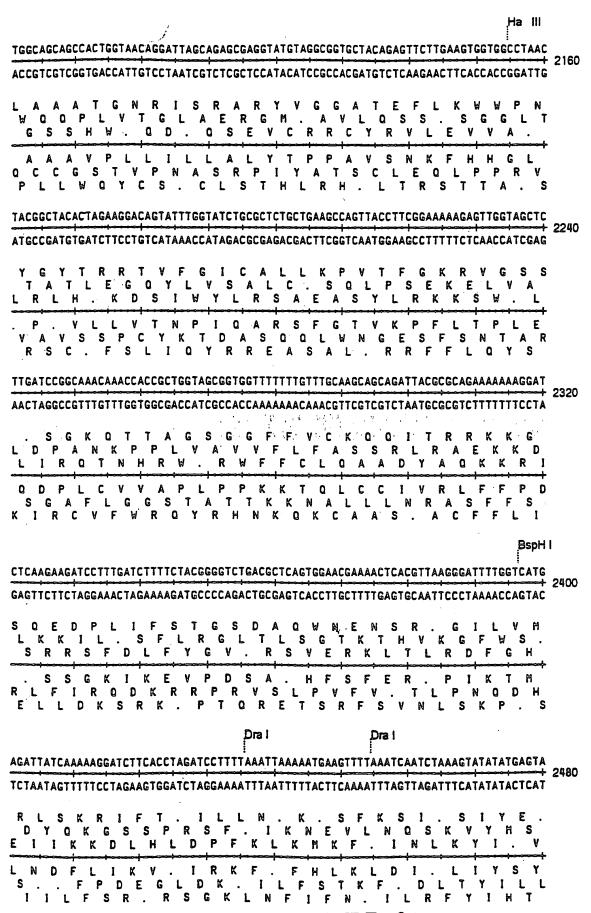
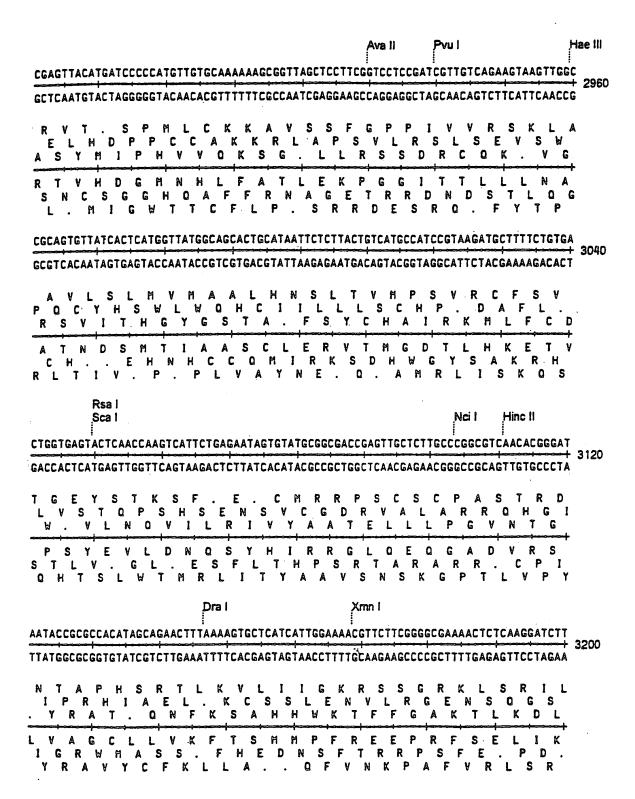


FIGURE 36

AACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTG + 2560 TTGAACCAGACTGTCAATGGTTACGAATTAGTCACTCCGTGGATAGAGTCGCTAGACAGATAAAGCAAGTAGGTATCAAC LISEAPISAI SVRHLSORS GTYLSD N C s c ILSAGIEAIO CRD. RLSRD CCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGAC + 2640 GGACTGAGGGGCAGCACATCTATTGATGCTATGCCCTCCCGAATGGTAGACCGGGGTCACGACGTTACTATGGCGCTCTG . LPVV. ITTIREGLPSGPSAAHIPRD PDSPSCR. LRYGRAYHLAPVLO YRF1 LTPRRVDNYDTGGLTIWPQCCNDTAR Q S G T T Y G S E G D H L IVVIRSPKGDPGLAA YSRYP ₩ R A T S CHYRSV R V G R R T S L . SVPPSVMQGWHQLSVALG Bgl I Hae III Ava II PR S P A P D L S A I N Q P A G R A E R R S G P A T L HAHRLOIYOO. TSOPEGPS/ TLTGSRFISNKPASRKGRA TSQPEGPSAE OKWSCNF G R E G A G S K D A I F W G A P L A S R L L P G A V K LWGSPGLASTTRCS LLLGALRF PRACFHDQL Asel. Nci I Fsp I ATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTG 2800 A S I Q S I N C C R E A R V S S S P V PSSLLIVAG KLE. VVRQLI PGS.SK.FAS QQRSALTL EGTL TRWNI SSYT IRRWGT.. NNGPL. LLYNAL. YNACRO SF G M ASF SSGSQRSR LLQAS A C HAR R L V # L H S AN D V V W Y G F I Q L R F TDRED N P AENL R K THS.EA G T

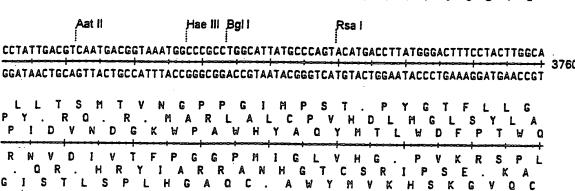


UM- UIJOUSKE

M2212041 1 -

ApaL I **ACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTT** + 3280 TGGCGACAACTCTAGGTCAAGCTACATTGGGTGAGCACGTGGGTTGACTAGAAGTCGTAGAAAATGAAAGTGGTCGCAAA LRSSSM.PTRAPN.SSASFT .DPVRCNPLVHPTDLQHLLL SP TAVEIOFD V THS C T Q L I F S I F Y F H Q R F Q D E A D K V K V L T S R . C R K S E G A N G S N L D L E I Y G V R A G L Q D E R O O S G T R H L G S T C G V S R . A G L V A T S I W N S T V W E H V W S I K L M K . CTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTC + 3360 GACCCACTCGTTTTTGTCCTTCCGTTTTACGGCGTTTTTTCCCTTATTCCCGCTGTGCCTTTACAACTTATGAGTATGAG S G . A K T G R O N A A K K G I R A T R K C . I L I L L G E O K O E G K M P O K R E . G R H G N V E Y S Y S W V S K N R K A K C R K K G N K G D T E M L N T H T PHAFVPLCFAAFFPILAVRFHQISMS RPSCFCSPLIGCFLSYPRCPFTSYEYE QTLLFLFAFHRLFPFLPSVSINFV.VF Hinc II Spe I Asel TICCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGCCGCGTTGACATTGATTATTGACTAGTTATTAA `` 3440 AAGGAAAAAGTTATAATAACTCGTAAATAGTCCCAATAACAGAGTACGCGCAACTGTAACTAATAACTGATCAATAATT FLFQYY. SIYQGYCLMRVDIDY. LVI SFFNIIEAFIRVIVSCALTLIID. LL PFSILLKHLSGLLSHAR. H. LLTSY. LVIN RK.Y.QLM..P.QRMRT EKKLIISANILTITEHAN GKEINNFCKDPNND.ARQ V N I I S . ARQCQNNVL Hae III Bgl I TAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCC + 3520 ATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCCATTTACCGGGCGG S N Q L R G H . F I A H I W S S A L H N L R . H A V I N Y G V I S S . P I Y G V P R Y I T Y G K W P . S I T G S L V H S P Y H E F R V T . L T V N G I

LL. NRP.. N M A W I H L E A N C L K R Y I A R T I L . P T M L E Y G M Y P T G R . M V . P L H G A Y D I V P D N T . L G Y I S N R T V Y S V T F P G G



RS8 | SNAB | Sty | RS8 |

GTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTTG

CATGTAGATGCATAATCAGTAGCGATAATGGTACCACTACGCCAAAACCGTCATGTAGTTACCCGCACCTATCGCCAAAC

3840

Nco I

BsaA I

S T S T Y . S S L L P W . C G F G S T S H G V D S G L V H L R I S H R Y Y H G D A V L A V H Q W A W I A V . Y I Y V L V I A I T H V H R F W Q Y I N G R G . R F V D V Y . D D S N G H H H P K P L V D I P T S L P K I C R R I L . R . . W P S A T K A T C . H A H I A T Q Y H . T N T H A I V H T I R N Q C Y H L P R P Y R N S

Aat II

THE DFQ VSTPLTSHEV CFGTKINGTFQ LTGISKSPPH. RQWEFVLAPKSTGLSDSRGFPSLHPIDVNGSLFWHQNQRDFPV.PSKWTEVGNVDIPTQKPVLILPVKWSVPIELDGGWQR. HSNTKAGFDVPSEL

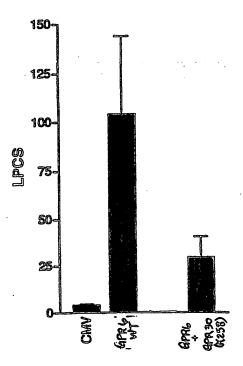
AAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCT
TTTACAGCATTGTTGAGGCGGGGTAACTGCGTTTACCCGCCCATCCGCACATGCCACCCTCCAGATATATTCGTCTCGAGA

N V V T T P P H . R K W A V G V Y G G R S I . A E L K H S . Q L R P I D A N G R . A C T V G G L Y K Q S S K C R N N S A P L T Q H G G R R V R W E V Y I S R A L F T T V V G G W Q R L H A T P T Y P P L D I Y A S S E I D Y C S R G H S A F P R Y A H V T P P R Y L C L E F H R L L E A G N V C I P P L R T R H S T . I L L A R

Asel

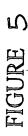
CTGGCTAACTAGAGAACCCACTGCTTAACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCC
GACCGATTGATCTCTTGGGTGACGAATTGACCGAATAGCTTTAATTATGCTGAGTGATATCCCTCTGGG

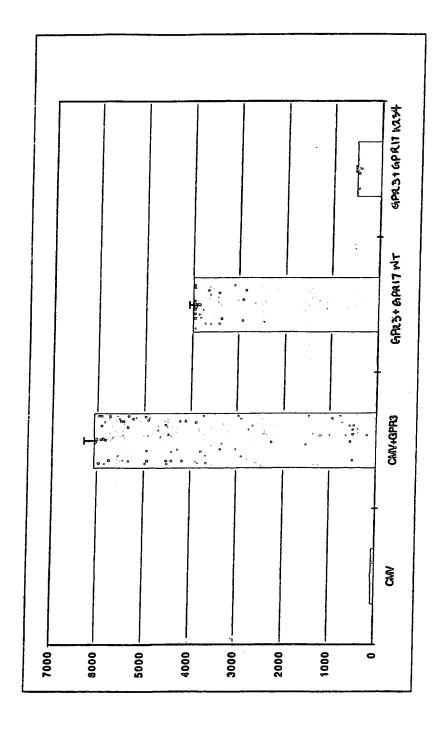
S G . L E N P L L N W L I E I N T T H Y R E T L A N . R T H C L T G L S K L I R L T I G R P W L T R E P T A . L A Y R N . Y D S L . G D P P . S S F G S S L Q S I S I L V V . . L S V W R A L . L V W Q K V P K D F N I R S V I P L G Q S V L S G V A . S A . R F . Y S E S Y P S G



Expression plamid

FIGURE 4





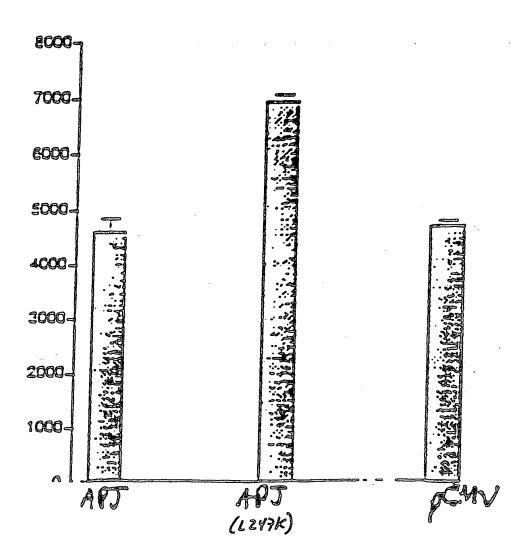


FIGURE 6

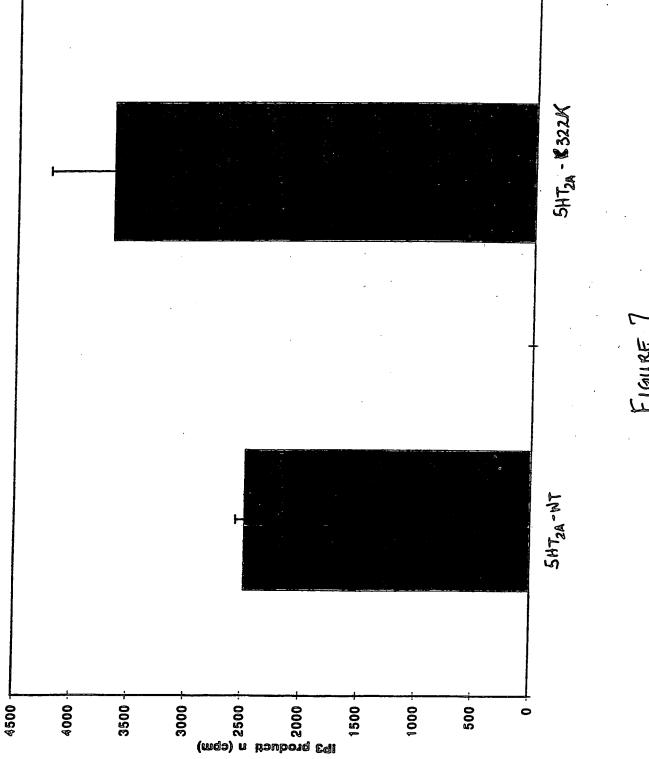


FIGURE 8A

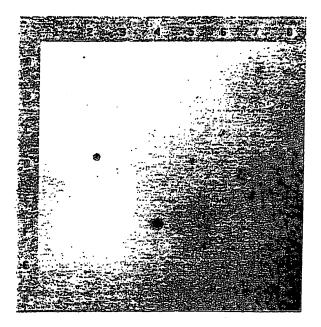
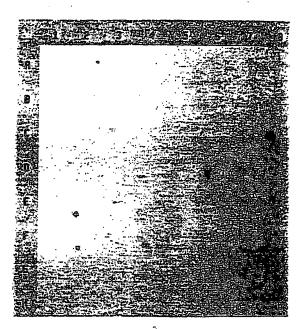


FIGURE 8B



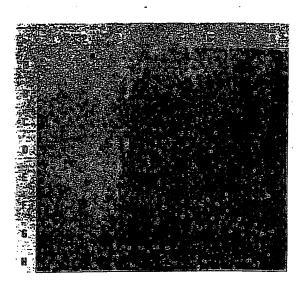


FIGURE 8C

SEQUENCE LISTING

(1) GENERAL INFORMATION: (i) APPLICANT: Behan, Dominic P. Chalmers, Derek T. 5 Liaw, Chen W. (ii) TITLE OF INVENTION: Non-Endogenous, Constitutively Activated Human G Protein-Coupled Orphan Receptors (iii) NUMBER OF SEQUENCES: 280 10 (iv) CORRESPONDENCE ADDRESS: (A) ADDRESSEE: Arena Pharmaceuticals, Inc. (B) STREET: 6166 Nancy Ridge Drive (C) CITY: San Diego 15 (D) STATE: CA (E) COUNTRY: USA (F) ZIP: 92122 (v) COMPUTER READABLE FORM: (A) MEDIUM TYPE: Floppy disk 20 (B) COMPUTER: IBM PC compatible (C) OPERATING SYSTEM: PC-DOS/MS-DOS (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (vi) CURRENT APPLICATION DATA: (A) APPLICATION NUMBER: US 25 (B) FILING DATE: (C) CLASSIFICATION: (viii) ATTORNEY/AGENT INFORMATION: (A) NAME: Burgoon, Richard P. (B) REGISTRATION NUMBER: 34,787 30 (ix) TELECOMMUNICATION INFORMATION: (A) TELEPHONE: (619) 453-7200 (B) TELEFAX: (619)453-7210 (2) INFORMATION FOR SEQ ID NO:1: (i) SEQUENCE CHARACTERISTICS: 35 (A) LENGTH: 1068 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATGGAAGATT TGGAGGAAAC ATTATTTGAA GAATTTGAAA ACTATTCCTA TGACCTAGAC

	TATTACTCTC	TGGAGTCTGA	TTTGGAGGAG	AAAGTCCAGC	TGGGAGTTGT	TCACTGGGTC	120
	TCCCTGGTGT	TATATTGTTT	GGCTTTTGTT	CTGGGAATTC	CAGGAAATGC	CATCGTCATT	180
	TGGTTCACGG	GGCTCAAGTG	GAAGAAGACA	GTCACCACTC	TGTGGTTCCT	CAATCTAGCC	240
	ATTGCGGATT	TCATTTTCT	TCTCTTTCTG	CCCCTGTACA	TCTCCTATGT	GGCCATGAAT	300
5	TTCCACTGGC	CCTTTGGCAT	CTGGCTGTGC	AAAGCCAATT	CCTTCACTGC	CCAGTTGAAC	360
	ATGTTTGCCA	GTGTTTTTT	CCTGACAGTG	ATCAGCCTGG	ACCACTATAT	CCACTTGATC	420
	CATCCTGTCT	TATCTCATCG	GCATCGAACC	CTCAAGAACT	CTCTGATTGT	CATTATATTC	480
	ATCTGGCTTT	TGGCTTCTCT	AATTGGCGGT	CCTGCCCTGT	ACTTCCGGGA	CACTGTGGAG	540
	TTCAATAATC	ATACTCTTTG	CTATAACAAT	TTTCAGAAGC	ATGATCCTGA	CCTCACTTTG	600
10	ATCAGGCACC	ATGTTCTGAC	TTGGGTGAAA	TTTATCATTG	GCTATCTCTT;	CCCTTTGCTA	660
	ACAATGAGTA	TTTGCTACTT	GTGTCTCATC	TTCAAGGTGA	AGAAGCGAAC	AGTCCTGATC	720
	TCCAGTAGGC	ATTTCTGGAC	AATTCTGGTT	GTGGTTGTGG	CCTTTGTGGT	TTGCTGGACT	780
	CCTTATCACC	TGTTTAGCAT	TTGGGAGCTC	ACCATTCACC	ACAATAGCTA	TTCCCACCAT	840
	GTGATGCAGG	CTGGAATCCC	CCTCTCCACT	GGTTTGGCAT	TCCTCAATAG	TTGCTTGAAC	900
15	CCCATCCTTT	ATGTCCTAAT	TAGTAAGAAG	TTCCAAGCTC	GCTTCCGGTC	CTCAGTTGCT	960
	GAGATACTCA	AGTACACACT	GTGGGAAGTC	AGCTGTTCTG	GCACAGTGAG	TGAACAGCTC	1020
	AGGAACTCAG	AAACCAAGAA	TCTGTGTCTC	CTGGAAACAG	CTCAATAA		1068
	/2\ TATEODAS	MTON BOD OF		•			

(3) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 355 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Glu Asp Leu Glu Glu Thr Leu Phe Glu Glu Phe Glu Asn Tyr Ser 1 5 10 15

Tyr Asp Leu Asp Tyr Tyr Ser Leu Glu Ser Asp Leu Glu Glu Lys Val

30 Gln Leu Gly Val Val His Trp Val Ser Leu Val Leu Tyr Cys Leu Ala
35 40 45

באופרטינות בואות המפוזמניאו : -

	Phe	Val 50	Leu	Gly	Ile	Pro	Gly 55	Asn	Ala	Ile	Val	Ile 60	Trp	Phe	Thr	Gl
	Leu 65	Lys	Trp	Lys	Lys	Thr 70	Val	Thr	Thr	Leu	Trp 75	Phe	Leu	Asn	Leu	Ala 80
5	Ile	Ala	Asp	Phe	Ile 85	Phe	Leu	Leu	Phe	Leu 90	Pro	Leu	Tyr	Ile	Ser 95	Ту
	Val	Ala	Met	Asn 100	Phe	His	Trp	Pro	Phe 105	Gly	Ile	Trp	Leu	Cys 110	Lys	Ala
10	Asn	Ser	Phe 115	Thr	Ala	Gln	Leu	Asn 120	Met	Phe	Ala	Ser	Val 125	Phe	Phe	Let
	Thr	Val 130	Ile	Ser	Leu	Asp	His 135	Tyr	Ile	His	Leu	Ile 140	His	Pro	Val	Leu
	Ser 145	His	Arg	His	Arg	Thr 150	Leu	Lys	Asn	Ser	Leu 155	Ile	Val	Ile	Ile	Phe 160
15 .	Ile	Trp	Leu	Leu	Ala 165	Ser	Leu	Ile	Gly	Gly 170	Pro	Ala	Leu	Tyr	Phe 175	Arg
	Asp	Thr	Val	Glu 180	Phe	Asn	Asn	His	Thr 185	Leu	Cys	Tyr	Asn	Asn 190	Phe	Glr
20	Lys	His	Asp 195	Pro	Asp	Leu	Thr	Leu 200	Ile	Arg	His	His	Val. 205	Leu	Thr	Trp
	Val	Lys 210	Phe	Ile	Ile	Gly	Tyr 215	Leu	Phe	Pro	Leu	Leu 220	Thr	Met	Ser	Ile
	Cys 225	Tyr	Leu	Cys	Leu	Ile 230	Phe	Lys	Val	Lys	Lys 235	Arg	Thr	Val	Leu	11e 240
25	Ser	Ser	Arg	His	Phe 245	Trp	Thr	Ile	Leu	Val 250	Val	Val	Val	Ala	Phe 255	Va]
	Val	Cys	Trp	Thr 260	Pro	Tyr			Phe 265	Ser	Ile	Trp	Glu	Leu 270	Thr	Ile
30			275					280					285		Pro	
		290					295					300			Leu	_
	305					310					315				Val	320
35	Glu	Ile	Leu	Lys	Tyr 325	Thr	Leu	Trp	Glu	Val 330	Ser	Cys	Ser	Gly	Thr 335	Va]
	Ser	Glu	Gln	Leu	Arq	Asn	Ser	Glu	Thr	Lvs	Asn	Leu	Cvs	Leu	Leu	Glu

340 345 350

Thr Ala Gln 355

(4) INFORMATION FOR SEQ ID NO:3:

- 5 (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1089 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- 10 (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATGGGCAACC ACACGTGGGA GGGCTGCCAC GTGGACTCGC GCGTGGACCA CCTCTTTCCG 60 CCATCCCTCT ACATCTTTGT CATCGGCGTG GGGCTGCCCA CCAACTGCCT GGCTCTGTGG 120 GCGGCCTACC GCCAGGTGCA ACAGCGCAAC GAGCTGGGCG TCTACCTGAT GAACCTCAGC 180 ATCGCCGACC TGCTGTACAT CTGCACGCTG CCGCTGTGGG TGGACTACTT CCTGCACCAC 240 GACAACTGGA TCCACGGCCC CGGGTCCTGC AAGCTCTTTG GGTTCATCTT CTACACCAAT 300 ATCTACATCA GCATCGCCTT CCTGTGCTGC ATCTCGGTGG ACCGCTACCT GGCTGTGGCC CACCCACTCC GCTTCGCCCG CCTGCGCCGC GTCAAGACCG CCGTGGCCGT GAGCTCCGTG 420 GTCTGGGCCA CGGAGCTGGG CGCCAACTCG GCGCCCCTGT TCCATGACGA GCTCTTCCGA 480 GACCGCTACA ACCACACCTT CTGCTTTGAG AAGTTCCCCA TGGAAGGCTG GGTGGCCTGG 540 ATGAACCTCT ATCGGGTGTT CGTGGGCTTC CTCTTCCCGT GGGCGCTCAT GCTGCTGTCG 600 TACCGGGGCA TCCTGCGGGC CGTGCGGGGC AGCGTGTCCA CCGAGCGCCA GGAGAAGGCC 660 AAGATCAAGC GGCTGGCCCT CAGCCTCATC GCCATCGTGC TGGTCTGCTT TGCGCCCTAT 720 CACGTGCTCT TGCTGTCCCG CAGCGCCATC TACCTGGGCC GCCCCTGGGA CTGCGGCTTC 780 25 GAGGAGCGCG TCTTTTCTGC ATACCACAGC TCACTGGCTT TCACCAGCCT CAACTGTGTG 840 GCGGACCCCA TCCTCTACTG CCTGGTCAAC GAGGGCGCCC GCAGCGATGT GGCCAAGGCC 900 CTGCACAACC TGCTCCGCTT TCTGGCCAGC GACAAGCCCC AGGAGATGGC CAATGCCTCG 960 CTCACCCTGG AGACCCCACT CACCTCCAAG AGGAACAGCA CAGCCAAAGC CATGACTGGC 1020 AGCTGGGCGG CCACTCCGCC TTCCCAGGGG GACCAGGTGC AGCTGAAGAT GCTGCCGCCA 1080 30 1089 GCACAATGA

(5) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 362 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein

	(x:	i) SI	EQUEI	NCE I	DESCI	RIPT	ION:	SEQ	ID 1	NO : 4	:					
10	Met 1	Gly	Asn	His	Thr 5	Trp	Glu	Gly	Cys	His 10	Val	Asp	Ser	Arg	Val 15	Asp
	His	Leu	Phe	Pro 20	Pro	Ser	Leu	Tyr	Ile 25	Phe	Val	Ile	Gly	Val	Gly	Leu
	Pro	Thr	Asn 35	Cys	Leu	Ala	Leu	Trp 40	Ala	Ala	Tyr	Arg	Gln 45	Val	Gln	Gln
15	Arg	Asn 50	Glu	Leu	Gly	Val	Tyr 55	Leu	Met	Asn	Leu	Ser 60	Ile	Ala	Asp	Leu
:	Leu 65	Tyr	Ile	Cys	Thr	Leu 70	Pro	Leu	Trp	Val	Asp 75	туг	Phe	Leu	His	His 80
20	Asp	Asn	Trp	Ile	His 85	Gly	Pro	Gly	Ser	Cys 90	Lys	Leu	Phe	Gly	Phe 95	Ile
·	Phe	Tyr	Thr	Asn 100	Ile	Tyr	Ile	Ser	Ile 105	Ala	Phe	Leu	Cys	Cys 110	Ile	Ser
	Val	Asp	Arg 115	Tyr	Leu	Ala	Val	Ala 120	His	Pro	Leu	Arg	Phe 125	Ala	Arg	Leu
25	Arg	Arg	Val	Lys	Thr	Ala	Val	Ala	Val	Ser	Ser	Val	Val	Trp	Ala	Thr

130 135 140

> Glu Leu Gly Ala Asn Ser Ala Pro Leu Phe His Asp Glu Leu Phe Arg 145

Asp Arg Tyr Asn His Thr Phe Cys Phe Glu Lys Phe Pro Met Glu Gly 30 170

> Trp Val Ala Trp Met Asn Leu Tyr Arg Val Phe Val Gly Phe Leu Phe 180 185 190

> Pro Trp Ala Leu Met Leu Leu Ser Tyr Arg Gly Ile Leu Arg Ala Val

35 Arg Gly Ser Val Ser Thr Glu Arg Gln Glu Lys Ala Lys Ile Lys Arg

Leu Ala Leu Ser Leu Ile Ala Ile Val Leu Val Cys Phe Ala Pro Tyr

	225		230		2	35		24	10
	His Val I	Leu Leu Leu 245	Ser Arg	Ser Ala	Ile T 250	yr Leu	Gly Arg	Pro T1 255	тр
5	Asp Cys G	Gly Phe Glu 260	Glu Arg	Val Phe 265		la Tyr	His Ser 270	Ser Le	eu
		Thr Ser Leu 275	Asn Cys	Val Ala 280	Asp P		Leu Tyr 285	Cys Le	eu
	Val Asn 0 290	Glu Gly Ala	Arg Ser 295		Ala L	ys Ala 300	Leu His	Asn Le	eu
10	Leu Arg I 305	Phe Leu Ala	Ser Asp	Lys Pro		Slu Met 315	Ala Asn		er 20
	Leu Thr I	Leu Glu Thr 325	Pro Leu	Thr Ser	Lys A	Arg Asn	Ser Thr	Ala Ly 335	γs
15	Ala Met 3	Thr Gly Ser 340	Trp Ala	Ala Thr 345		Pro Ser	Gln Gly 350	Asp G	ln
		Leu Lys Met 355	Leu Pro	Pro Ala 360	Gln				
	(6) INFORMATIO	ON FOR SEQ	ID NO:5:					. •	
20	(A) (B) (C)	UENCE CHARAG LENGTH: 30 TYPE: nucle STRANDEDNES TOPOLOGY: 3	base pa eic acid SS: sing	irs					
	(ii) MO	LECULE TYPE	: DNA (g	enomic)					
25	(xi) SE	QUENCE DESC	RIPTION:	SEQ ID	NO:5:			•	
	TATGAATTCA GA	TGCTCTAA AC	GTCCCTGC	!					30
	(7) INFORMATION	ON FOR SEQ	ID NO:6:						
30	(A) (B) (C)	UENCE CHARA LENGTH: 30 TYPE: nucl STRANDEDNE TOPOLOGY:	base pa eic acid SS: sing	irs l					
	(ii) MO	LECULE TYPE	: DNA (g	genomic)					
	(xi) SE	QUENCE DESC	RIPTION:	SEQ ID	NO:6:	-			
35	TCCGGATCCA CC	CTGCACCTG CG	CCTGCAC	2					30

(8) INFORMATION FOR SEQ ID NO: 7:

WO 00/22129 PCT/US99/23938

7

(i)	SEQUENCE	CHARACTERISTICS:

- (A) LENGTH: 1002 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- 5 (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ATGGAGTCCT CAGGCAACCC AGAGAGCACC ACCTTTTTTT ACTATGACCT TCAGAGCCAG 60 CCGTGTGAGA ACCAGGCCTG GGTCTTTGCT ACCCTCGCCA CCACTGTCCT GTACTGCCTG 120 10 GTGTTTCTCC TCAGCCTAGT GGGCAACAGC CTGGTCCTGT GGGTCCTGGT GAAGTATGAG 180 AGCCTGGAGT CCCTCACCAA CATCTTCATC CTCAACCTGT GCCTCTCAGA CCTGGTGTTC 240 GCCTGCTTGT TGCCTGTGTG GATCTCCCCA TACCACTGGG GCTGGGTGCT GGGAGACTTC 300 CTCTGCAAAC TCCTCAATAT GATCTTCTCC ATCAGCCTCT ACAGCAGCAT CTTCTTCCTG 360 ACCATCATGA CCATCCACCG CTACCTGTCG GTAGTGAGCC CCCTCTCCAC CCTGCGCGTC 420 15 CCCACCTCC GCTGCCGGGT GCTGGTGACC ATGGCTGTGT GGGTAGCCAG CATCCTGTCC 480 TCCATCCTCG ACACCATCTT CCACAAGGTG CTTTCTTCGG GCTGTGATTA TTCCGAACTC 540 ACGTGGTACC TCACCTCGT CTACCAGCAC AACCTCTTCT TCCTGCTGTC CCTGGGGATT 600 ATCCTGTTCT GCTACGTGGA GATCCTCAGG ACCCTGTTCC GCTCACGCTC CAAGCGGCGC 660 CACCGCACGG TCAAGCTCAT CTTCGCCATC GTGGTGGCCT ACTTCCTCAG CTGGGGTCCC 720 20 TACAACTTCA CCCTGTTTCT GCAGACGCTG TTTCGGACCC AGATCATCCG GAGCTGCGAG 780 GCCAAACAGC AGCTAGAATA CGCCCTGCTC ATCTGCCGCA ACCTCGCCTT CTCCCACTGC 840 TGCTTTAACC CGGTGCTCTA TGTCTTCGTG GGGGTCAAGT TCCGCACACA CCTGAAACAT 900 GTTCTCCGGC AGTTCTGGTT CTGCCGGCTG CAGGCACCCA GCCCAGCCTC GATCCCCCAC 960 TCCCCTGGTG CCTTCGCCTA TGAGGGCGCC TCCTTCTACT GA 1002

- 25 (9) INFORMATION FOR SEQ ID NO:8:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 333 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
- 30 (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

	Met 1	Glu	Ser	Ser	Gly 5	Asn	Pro	Glu	Ser	Thr 10	Thr	Phe	Phe	Tyr	Tyr 15	Asp
	Leu	Gln	Ser	Gln 20	Pro	Cys	Glu	Asn	Gln 25	Ala	Trp	Val	Phe	Ala 30	Thr	Leu
5	Ala	Thr	Thr 35	Val	Leu	Tyr	Cys	Leu 40	Val	Phe	Leu	Leu	Ser 45	Leu	Val	Gly
	Asn	Ser 50	Leu	Val	Leu	Trp	Val 55	Leu	Val	Lys	Tyr	Glu 60	Ser	Leu	Glu	Ser
10	Leu 65	Thr	Asn	Ile	Phe	Ile 70	Leu	Asn	Leu	Cys	Leu 75	Ser	Asp	Leu	Val	Phe 80
	Ala	Cys	Leu	Leu	Pro 85	Val	Trp	Ile	Ser	Pro 90	Tyr	His	Trp	Gly	Trp 95	Val
	Leu	Gly	Asp	Phe 100	Leu	Cys	Lys	Leu	Leu 105	Asn	Met	Ile	Phe	Ser 110	Ile	Ser
15	Leu	Tyr	Ser 115	Ser	Ile	Phe	Phe	Leu 120	Thr	Ile	Met	Thr	Ile 125	His	Arg	Tyr
	Leu	Ser 130	Val	Val	Ser	Pro	Leu 135	Ser	Thr	Leu	Arg	Val 140	Pro	Thr	Leu	Arg
20	Cys 145	Arg	Val	Leu	Val	Thr 150	Met	Ala	Val	Trp	Val 155	Ala	Ser	Ile	Leu	Ser 160
	Ser	Ile	Leu	Asp	Thr 165	Ile	Phe	His	Lys	Val 170	Leu	Ser	Ser	Gly	Cys 175	Asp
	Tyr	Ser	Glu	Leu 180	Thr	Trp	Tyr	Leu	Thr 185	Ser	Val	Tyr	Gln	His 190	Asn	Leu
25	Phe	Phe	Leu 195	Leu	Ser	Leu	Gly	Ile 200	Ile	Leu	Phe	Cys	Tyr 205	Val	Glu	Ile
·	Leu	Arg 210	Thr	Leu	Phe	Arg	Ser 215	Arg	Ser	Lys	Arg	Arg 220	His	Arg	Thr	Val
30	Lys 225	Leu	Ile	Phe	Ala	Ile 230	Val	Val	Ala	Tyr	Phe 235	Leu	Ser	Trp	Gly	Pro 240
	Tyr	Asn	Phe	Thr	Leu 245	Phe	Leu	Gln	Thr	Leu 250	Phe	Arg	Thr	Gln	Ile 255	Ile
	Arg	Ser	Cys	Glu 260	Ala	Lys	Gln	Gln	Leu 265	Glu	Tyr	Ala	Leu	Leu 270	Ile	Cys
35	Arg	Asn	Leu 275	Ala	Phe	Ser	His	Cys 280	Cys	Phe	Asn	Pro	Val 285	Leu	Tyr	Val

	Phe Val Gly Val Lys Phe Arg Thr His Leu Lys His Val Leu Arg Gln 290 295 300	
	Phe Trp Phe Cys Arg Leu Gln Ala Pro Ser Pro Ala Ser Ile Pro His 305 310 315 320	
5	Ser Pro Gly Ala Phe Ala Tyr Glu Gly Ala Ser Phe Tyr 325 330	
	(10) INFORMATION FOR SEQ ID NO:9:	
10	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:	
15	GCAAGCTTGG GGGACGCCAG GTCGCCGGCT	30
	(11) INFORMATION FOR SEQ ID NO:10:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid	
20	(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:	
	GCGGATCCGG ACGCTGGGGG AGTCAGGCTG C	31
25	(12) INFORMATION FOR SEQ ID NO:11:	
	(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 987 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single	
30	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:	
	ATGGACAACG CCTCGTTCTC GGAGCCCTGG CCCGCCAACG CATCGGGCCC GGACCCGGCG	60
	CTGAGCTGCT CCAACGCGTC GACTCTGGCG CCGCTGCCGG CGCCGCTGGC GGTGGCTGTA	L20
35	CCAGTTGTCT ACGCGGTGAT CTGCGCCGTG GGTCTGGCGG GCAACTCCGC CGTGCTGTAC	180

GTGTTGCTGC GGGCGCCCCG CATGAAGACC GTCACCAACC TGTTCATCCT CAACCTGGCC

	ATCGCCGACG AGCTCTTCAC GCTGGTGCTG CCCATCAACA TCGCCGACTT CCTGCTGCGG	300
	CAGTGGCCCT TCGGGGAGCT CATGTGCAAG CTCATCGTGG CTATCGACCA GTACAACACC	360
	TTCTCCAGCC TCTACTTCCT CACCGTCATG AGCGCCGACC GCTACCTGGT GGTGTTGGCC	420
5	ACTGCGGAGT CGCGCCGGGT GGCCGGCCGC ACCTACAGCG CCGCGCGCGC GGTGAGCCTG	480
	GCCGTGTGGG GGATCGTCAC ACTCGTCGTG CTGCCCTTCG CAGTCTTCGC CCGGCTAGAC	540
	GACGAGCAGG GCCGGCGCA GTGCGTGCTA GTCTTTCCGC AGCCCGAGGC CTTCTGGTGG	600
	CGCGCGAGCC GCCTCTACAC GCTCGTGCTG GGCTTCGCCA TCCCCGTGTC CACCATCTGT	660
	GTCCTCTATA CCACCCTGCT GTGCCGGCTG CATGCCATGC	720
10	GCCCTGGAGC GCGCCAAGAA GCGGGTGACC TTCCTGGTGG TGGCAATCCT GGCGGTGTGC	780
	CTCCTCTGCT GGACGCCCTA CCACCTGAGC ACCGTGGTGG CGCTCACCAC CGACCTCCCG	840
	CAGACGCCGC TGGTCATCGC TATCTCCTAC TTCATCACCA GCCTGACGTA CGCCAACAGC	900
	TGCCTCAACC CCTTCCTCTA CGCCTTCCTG GACGCCAGCT TCCGCAGGAA CCTCCGCCAG	960
	CTGATAACTT GCCGCGGGC AGCCTGA	987
15	(13) INFORMATION FOR SEQ ID NO:12:	
20	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 328 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: (D) TOPOLOGY: not relevant 	
	(ii) MOLECULE TYPE: protein	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:	
	Met Asp Asn Ala Ser Phe Ser Glu Pro Trp Pro Ala Asn Ala Ser Gly 1 5 10 15	•
25	Pro Asp Pro Ala Leu Ser Cys Ser Asn Ala Ser Thr Leu Ala Pro Leu 20 25 30	ı .*
	Pro Ala Pro Leu Ala Val Ala Val Pro Val Val Tyr Ala Val Ile Cys 35 40 45	;
30	Ala Val Gly Leu Ala Gly Asn Ser Ala Val Leu Tyr Val Leu Leu Arg 50 55 60	ſ
	Ala Pro Arg Met Lys Thr Val Thr Asn Leu Phe Ile Leu Asn Leu Ala	L

ONEDDOID AND CONSTRUCT

		Ile	Ala	Asp	Glu	Leu 85	Phe	Thr	Leu	Val	Leù 90	Pro	Ile	Asn	Ile	Ala 95	Asp
		Phe	Leu	Leu	Arg 100	Gln	Trp	Pro	Phe	Gly 105	Glu	Leu	Met	Cys	Lys 110	Leu	Ile
5		Val	Ala	Ile 115	Asp	Gln	Tyr	Asn	Thr 120	Phe	Ser	Ser	Leu	Tyr 125	Phe	Leu	Thr
		Val	Met 130	Ser	Ala	Asp	Arg	Tyr 135	Leu	Val	Val	Leu	Ala 140	Thr	Ala	Glu	Ser
10		Arg 145	Arg	Val	Ala	Gly	Arg 150	Thr	Tyr	Ser	Ala	Ala 155	Arg	Ala	Val	Ser	Leu 160
	,	Ala	Val	Trp	Gly	Ile 165	Val	Thr	Leu	Val	Val 170	Leu	Pro	Phe	Ala	Val 175	Phe
		Ala	Arg	Leu	Asp 180	Asp	Glu	Gln	Gly	Arg 185	Arg	Gln	Cys	Val	Leu 190	Val	Phe
15		Pro	Gln	Pro 195	Glu	Ala	Phe	Trp	Trp 200	Arg	Ala	Ser	Arg	Leu 205	Tyr	Thr	Leu
•		Val	Leu 210	Gly	Phe	Ala	Ile	Pro 215	Val	Ser	Thr	Ile	Cys 220	Val	Leu	Tyr	Thr
20		Thr 225	Leu	Leu	Cys	Arg	Leu 230	His	Ala	Met	Arg	Leu 235	Asp	Ser	His	Ala	Lys 240
		Ala	Leu	Glu	Arg	Ala 245	Lys	Lys	Arg	Val	Thr 250	Phe	Leu	Val	Val	Ala 255	Ile
		Leu	Ala	Val	Cys 260	Leu	Leu	Cys	Trp	Thr 265	Pro	Tyr	His	Leu	Ser 270	Thr	Val
25		Val	Ala	Leu 275	Thr	Thr	Asp	Leu	Pro 280	Gln	Thr	Pro	Leu	Val 285	Ile	Ala	Ile
		Ser	Tyr 290	Phe	Ile	Thr	Ser	Leu 295	Thr	Tyr	Ala	Asn	Ser 300	Cys ·	Leu	Asn	Pro
30		Phe 305	Leu	Tyr	Ala	Phe	Leu 310	Asp	Ala	Ser	Phe	Arg 315	Arg	Asn	Leu	Arg	Gln 320
		Leu	Ile	Thr	Cys	Arg 325	Ala	Ala	Ala								
	(14)	INFO	RMAT	CION	FOR	SEQ	ID N	10:13	:								

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:	
	CGGAATTCGT CAACGGTCCC AGCTACAATG	30
	(15) INFORMATION FOR SEQ ID NO:14:	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
10	(ii) MOLECULE TYPE: DNA (genomic) (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:	
:	ATGGATCCCA GGCCCTTCAG CACCGCAATA T	31
	(16) INFORMATION FOR SEQ ID NO:15:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1002 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:	
	ATGCAGGCCG CTGGGCACCC AGAGCCCCTT GACAGCAGGG GCTCCTTCTC CCTCCCCACG	60
	ATGGGTGCCA ACGTCTCTCA GGACAATGGC ACTGGCCACA ATGCCACCTT CTCCGAGCCA	120
	CTGCCGTTCC TCTATGTGCT CCTGCCCGCC GTGTACTCCG GGATCTGTGC TGTGGGGCTG	180
	ACTGGCAACA CGGCCGTCAT CCTTGTAATC CTAAGGGCGC CCAAGATGAA GACGGTGACC	240
25	AACGTGTTCA TCCTGAACCT GGCCGTCGCC GACGGGCTCT TCACGCTGGT ACTGCCCGTC	300
	AACATCGCGG AGCACCTGCT GCAGTACTGG CCCTTCGGGG AGCTGCTCTG CAAGCTGGTG	360
	CTGGCCGTCG ACCACTACAA CATCTTCTCC AGCATCTACT TCCTAGCCGT GATGAGCGTG	420
	GACCGATACC TGGTGGTGCT GGCCACCGTG AGGTCCCGCC ACATGCCCTG GCGCACCTAC	480
	CGGGGGGCGA AGGTCGCCAG CCTGTGTGTC TGGCTGGGCG TCACGGTCCT GGTTCTGCCC	540
30	TTCTTCTCTT TCGCTGGCGT CTACAGCAAC GAGCTGCAGG TCCCAAGCTG TGGGCTGAGC	600
	TTCCCGTGGC CCGAGCGGGT CTGGTTCAAG GCCAGCCGTG TCTACACTTT GGTCCTGGGC	660
	TTCGTGCTGC CCGTGTGCAC CATCTGTGTG CTCTACACAG ACCTCCTGCG CAGGCTGCGG	720

	GCCGTGCG	GC T	CCGC	TCTG	G AG	CCAA	GGCT	CTA	GGCA	AGG	CCAG	GCGG	AA G	GTGA	CCGT	С	780
	CTGGTCCT	CG T	CGTG	CTGG	C CG	TGTG	CCTC	CTC	TGCT	GGA	CGCC	CTTC	CA C	CTGG	CCTC	т	840
	GTCGTGGC	CC T	GACC.	acge	A CC	TGCC	CCAG	ACC	CCAC	TGG	TCAT	CAGT	AT G	TCCT.	ACGT	C	900
	ATCACCAG	CC T	CACG'	TACG	C CA	ACTC	GTGC	CTG	AACC	CĊT	TCCT	CTAC	gë c	TTTC	TAGA'	T · ·	960
5	GACAACTT	CC G	GAAG	AACT	T CC	GCAG	CATA	TTG	CGGT	GCT	GA					1	002
	(17) INF	ORMA	TION	FOR	SEQ	ID 1	NO:1	6 :				,					
10		(B (C (D) LEI) TYI) STI) TOI	NGTH PE: RAND: POLO	: 33: amine EDNE: GY: 1	3 am 5 ac SS: not:	ino a id rele	acid vant	s								
	(1	i) M	OLEC	OLE '	TYPE	: pro	otei	n.							:	•	
	(x	i) S	EQUEI	NCE I	DESCI	RIPT	ION:	SEQ	ID I	NO:1	6:						
15	Met 1	Gln	Ala	Ala	Gly 5	His	Pro	Glu	Pro	Leu 10	Asp	Ser	Arg	Gly	Ser 15	Phe	
	Ser	Leu	Pro	Thr 20	Met	Gly	Ala	Asn	Val 25	Ser	Gln	Asp	Asn	Gly 30	Thr	Gly	
	His	Asn	Ala 35	Thr	Phe	Ser		Pro 40	Leu	Pro	Phe	Leu	Tyr 45	Val	Leu	Leu	,
20	Pro	Ala 50	Val	Tyr	Ser	Gly	Ile 55	Cys	Ala	Val	Gly	Leu 60	Thr	Gly	Asn	Thr	· (.
	Ala 65	Val	Ile	Leu		Ile 70	Leu	Arg	Ala	Pro	Lys 75	Met	Lys	Thr	Val	Thr 80	
25	Asn	Val	Phe	Ile	Leu 85	Asn	Leu	Ala	Val	Ala 90	Asp	Gly	Leu	Phe	Thr 95	Leu	
	Val	Leu	Pro	Val	Asn	Ile	Ala	Glu	His 105	Leu	Leu	Gln	Tyr	Trp 110	Pro	Phe	•
	Gly	Glu	Leu 115	Leu	Cys	Lys	Leu	Val 120	Leu	Ala	Val	Asp	His 125	Tyr	Asn	Ile	
30	Phe	Ser 130	Ser	Ile	Tyr	Phe	Leu 135	Ala	Val	Met	Ser	Val 140	Asp	Arg	Tyr	Leu	
	Val	Val	Leu	Ala	Thr	Val 150	Arg	Ser	Arg	His	Met 155	Pro	Trp	Arg	Thr	Tyr 160	
35	Arg	Gly	Ala	Lys	Val 165	Ala	Ser	Leu	Cys	Val 170	Trp	Leu	Gly	Val	Thr 175	Val	

		1	Leu	Val	Leu	Pro 180	Phe	Phe	Ser	Phe	Ala 185	Gly	Val	Tyr	Ser	Asn 190	Glu	Leu	
		C	Sln	Val	Pro 195	Ser	Cys	Gly	Leu	Ser 200	Phe	Pro	Trp	Pro	Glu 205	Arg	.Val	Trp	
	5	I	Phe	Lys 210	Ala	Ser	Arg	Val	Tyr 215	Thr	Leu	Val	Leu	Gly 220	Phe	Val	Leu	Pro	
			Val 225	Cys	Thr	Ile	Cys	Val 230	Leu	Tyr	Thr	Asp	Leu 235	Leu	Arg	Arg	Leu	Arg 240	
	10	1	Ala	Val	Arg	Leu	Arg 245	Ser	Gly	Ala	Lys	Ala 250	Leu	Gly	Lys	Ala	Arg 255	Arg	
		1	Lys	Val	Thr	Val 260	Leu	Val	Leu	Val	Val 265	Leu	Ala	Val	Cys	Leu 270	Leu	Cys	
			Trp	Thr	Pro 275	Phe	His	Leu	Ala	Ser 280	Val	Val	Ala	Leu	Thr 285	Thr	Asp	Leu	
	15	1	Pro	Gln 290	Thr	Pro	Leu	Val	Ile 295	Ser	Met	Ser	Tyr	Val 300	Ile	Thr	Ser	Leu	
٠.			Thr 305	-	Ala		Ser	Cys 310		Asn		Phe		Tyr	Ala	Phe	Leu	Asp 320	
	20	• • ;	Asp	Asn	Phe	Arg	Lys 325	Asn	Phe	Arg	Ser	Ile 330	Leu	Arg	Cys	• •	•		
		(18)	INFO	ORMA!	rion	FOR	SEQ	ID I	NO:1	7:									
			(i)			CE C											•		
	25			(B)	TY:	PE:	nucl	eic :	acid										
						POLO										-			
			(i:	i) M	OLEC	ULE '	TYPE	: DN	A (g	enom	ic)								
			(x:	i) S	EQUE	NCE :	DESC	RIPT	ION:	SEQ	ID:	NO:1	7:						
		ACGAA	TTC	AG C	CATG	GTCC	T TG	AGGT	GAGT	GAC	CACC	AAG	TGCT.	TAAA					48
	30	(19)	INF	ORMA'	TION	FOR	SEQ	ID	NO:1	8:									
		-	(i)	(A (B) LE) TY	CE C NGTH PE:	: 27 nucl	bas eic	e pa acid	irs									
	35	٠				RAND POLO				le				,					
			(i	i) M	OLEC	ULE	TYPE	: DN	A (g	enom	ic)								
			1~	41 G	FOITE	NCE	חביפר	דם ד סי	TON:	SEC	ID	NO:1	8:						

WO 00/22129 PCT/US99/23938

GAGGATCCTG GAATGCGGGG AAGTCAG	27													
(20) INFORMATION FOR SEQ ID NO:19:														
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1107 base pairs (B) TYPE: nucleic acid														
(C) STRANDEDNESS: single (D) TOPOLOGY: linear														
(ii) MOLECULE TYPE: DNA (genomic)														
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:														
ATGGTCCTTG AGGTGAGTGA CCACCAAGTG CTAAATGACG CCGAGGTTGC CGCCCTCCTG	60													
GAGAACTTCA GCTCTTCCTA TGACTATGGA GAAAACGAGA GTGACTCGTG CTGTACCTCC	120													
CCGCCCTGCC CACAGGACTT CAGCCTGAAC TTCGACCGGG CCTTCCTGCC AGCCCTCTAC	180													
AGCCTCCTCT TTCTGCTGGG GCTGCTGGGC AACGGCGCGG TGGCAGCCGT GCTGCTGAGC	240													
CGGCGGACAG CCCTGAGCAG CACCGACACC TTCCTGCTCC ACCTAGCTGT AGCAGACACG	300													
CTGCTGGTGC TGACACTGCC GCTCTGGGCA GTGGACGCTG CCGTCCAGTG GGTCTTTGGC	360													
TCTGGCCTCT GCAAAGTGGC AGGTGCCCTC TTCAACATCA ACTTCTACGC AGGAGCCCTC	420													
CTGCTGGCCT GCATCAGCTT TGACCGCTAC CTGAACATAG TTCATGCCAC CCAGCTCTAC	480													
CGCCGGGGGC CCCCGGCCCG CGTGACCCTC ACCTGCCTGG CTGTCTGGGG GCTCTGCCTG	540													
CTTTTCGCCC TCCCAGACTT CATCTTCCTG TCGGCCCACC ACGACGAGCG CCTCAACGCC	600													
ACCCACTGCC AATACAACTT CCCACAGGTG GGCCGCACGG CTCTGCGGGT GCTGCAGCTG	660													
GTGGCTGGCT TTCTGCTGCC CCTGCTGGTC ATGGCCTACT GCTATGCCCA CATCCTGGCC	720													
GTGCTGCTGG TTTCCAGGGG CCAGCGGCGC CTGCGGGCCA TGCGGCTGGT GGTGGTGGTC	780													
GTGGTGGCCT TTGCCCTCTG CTGGACCCCC TATCACCTGG TGGTGCTGGT GGACATCCTC	840													
ATGGACCTGG GCGCTTTGGC CCGCAACTGT GGCCGAGAAA GCAGGGTAGA CGTGGCCAAG	900													
TCGGTCACCT CAGGCCTGGG CTACATGCAC TGCTGCCTCA ACCCGCTGCT CTATGCCTTT	960													
GTAGGGGTCA AGTTCCGGGA GCGGATGTGG ATGCTGCTCT TGCGCCTGGG CTGCCCCAAC	1020													
CAGAGAGGCC TCCAGAGGCA GCCATCGTCT TCCCGCCGGG ATTCATCCTG GTCTGAGACC	1080													
TONGROOCE CUTACTOCCC CUTCTCA	3305													

(21) INFORMATION FOR SEQ ID NO:20:

5	(A) LENGTH: 368 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: (D) TOPOLOGY: not relevant															
• .	(ii	.) -MC	LECU	LE 1	YPE:	pro	oteir	ı ·	,	•				·. :		
	(xi) SE	EQUEN	ICE I	ESCR	RIPTI	ON:	SEQ	ID N	10:20):					
	Met 1	Val	Leu	Glu	Val 5	Ser	Asp	His	Gln	Val 10	Leu	Asn	Asp	Ala	Glu 15	Val
10	Ala	Ala	Leu	Leu 20	Glu	Asn	Phe	Ser	Ser 25	Ser	Tyr	Asp	Tyr	Gly 30	Glu	Asn
٠.	Glu	Ser	Asp 35	Ser	Cys	Cys	Thr	Ser 40	Pro	Pro	Cys	Pro	Gln 45	Asp	Phe	Ser
15	Leu	Asn 50	Phe	Asp	Arg	Ala	Phe 55	Leu	Pro	Ala	Leu	Tyr 60	Ser	Leu	Leu	Phe
	Leu 65	Leu	Gly	Leu	Leu	Gly 70	Asn	Gly	Ala	Val	Ala 75	Ala	Val	Leu	Leu	Ser 80
	Arg	Arg	Thr	Ala	Leu 85	Ser	Ser	Thr	Asp	Thr 90	Phe	Leu	Leu	His	Leu 95	Ala
20	Val	Ala	Asp	Thr 100	Leu	Leu !	Val	Leu	Thr 105	Leu	Pro	Leu	Trp	Ala 110	Val	Asp
	Ala	Ala	Val 115	Gln	Trp	Val	Phe	Gly 120	Ser	Gly	Leu	Cys	Lys 125	Val	Ala	Gly
25	Ala	Leu 130	Phe	Asn	Ile	Asn	Phe 135	Tyr	Ala	Gly	Ala	Leu 140	Leu	Leu	Ala	Cys
	Ile 145	Ser	Phe	Asp	Arg	Tyr 150	Leu	Asn	Ile	Val	His 155	Ala	Thr	Gln	Leu	Tyr 160
	Arg	Arg	Gly	Pro	Pro 165	Ala	Arg	Val	Thr	Leu 170	Thr	Cys	Leu	Ala	Val 175	Trp
30	Gly	Leu	Cys	Leu 180	Leu	Phe	Ala	Leu	Pro 185	Asp	Phe	Ile	Phe	Leu 190	Ser	Ala
	His	His	Asp 195	Glu	Arg	Leu	Asn	Ala 200		His	Cys	Gln	Tyr 205	Asn	Phe	Pro
35	Gln	Val 210	Gly	Arg	Thr	Ala	Leu 215		Val	Leu	Gln	Leu 220		Ala	Gly	Phe
	Leu 225		Pro	Leu	Leu	Val 230		Ala	Tyr	Cys	Tyr 235	Ala	His	Ile	Leu	Ala 240

	Val	Leu	Leu	Val	Ser 245	Arg	Gly	Gln	Arg	Arg 250	Leu	Arg	Ala	Met	Arg 255	Leu
	Val	Val	Val	Val 260	Val	Val	Ala	Phe	Ala 265	Leu	Cys	Trp	Thr	Pro 270	Tyr	His
5	Leu	Val	Val 275	Leu	Val	Asp	Ile	Leu 280	Met	Asp	Leu	Gly	Ala 285	Leu	Ala	Arg
	Asn	Cys 290	-	Arg	Glu	Ser	Arg 295		Asp	Val	Ala	Lys 300	Ser	Val	Thr	Ser
10	Gly 305	Leu	Gly	Tyr	Met	His 310		Cys	Leu	Asn	Pro 315	Leu	Leu	Tyr	Ala	Phe 320
	Val	Gly	Val	Lys	Phe 325	Arg	Glu	Arg	Met	Trp 330	Met	Leu	Leu	Leu	Arg 335	Leu
	Gly	Cys	Pro	Asn 340	Gln	Arg	Gly	Leu	Gln 345	Arg	Gln	Pro	Ser	Ser 350	Ser	Arg
	Arg	Asp	Ser 355	Ser	Trp	Ser	Glu	Thr 360	Ser	Glu	Ala	Ser	Tyr 365	Ser	Gly	Leu
	(22) IN	FORM	OITA	1 FOR	SEÇ	, ID	NO:2	21:				<i>i</i> .		* **		lat
20	(i)	(B)	QUENC LEN TYP STF	NGTH: PE: r RANDE	: 30 nucle	base ic a	e pai acid singl	irs								
	(3.3	L) MC		:		:		· PDOM	 (c)	•			-			
		i) SI								1 0:21	L:					
25	TTAAGCTTC		٠													30
	(23) INFO	RMAT	rion	FOR	SEQ	ID 1	10:22	2:								
	· (i)	SEÇ	QUENC													
30		(B)	TYI STF	PE: r RANDE	nucle EDNES	eic a SS: s	acid sing]		٠.							
	(ii	i) MC	DLECU	JLE 1	YPE:	DN2	A (ge	enom	ic)							
	(xi	i) SI	EQUE	ICE I	ESCF	RIPT	ON:	SEQ	ID 1	10:22	2:					
	TTGGATCC	AA AA	AGAAC	CATO	G CAC	CTC	AGAG					• • •				30
35	(24) INFO	ORMAT	rion	FOR	SEQ	ID 1	NO:23	3:								
	431	SECT	TENCE	CUI	רים מסו	יד פיפי	בידרים	2.								

- (A) LENGTH: 1074 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- 5 (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

	ATGGCTGATG	ACTATGGCTC	TGAATCCACA	TCTTCCATGG	AAGACTACGT	TAACTTCAAC	60
	TTCACTGACT	TCTACTGTGA	GAAAAACAAT	GTCAGGCAGT	TTGCGAGCCA	TTTCCTCCCA	120
	CCCTTGTACT	GGCTCGTGTT	CATCGTGGGT	GCCTTGGGCA	ACAGTCTTGT	TATCCTTGTC	180
10	TACTGGTACT	GCACAAGAGT	GAAGACCATG	ACCGACATGT	TCCTTTTGAA	TTTGGCAATT	240
	GCTGACCTCC	TCTTTCTTGT	CACTCTTCCC	TTCTGGGCCA	TTGCTGCTGC	TGACCAGTGG	300
	AAGTTCCAGA	CCTTCATGTG	CAAGGTGGTC	AACAGCATGT	ACAAGATGAA	CTTCTACAGC	360
	TGTGTGTTGC	TGATCATGTG	CATCAGCGTG	GACAGGTACA	TTGCCATTGC	CCAGGCCATG	420
	AGAGCACATA	CTTGGAGGGA	GAAAAGGCTT	TTGTACAGCA	AAATGGTTTG	CTTTACCATC	480
15	TGGGTATTGG	CAGCTGCTCT	CTGCATCCCA	GAAATCTTAT	ACAGCCAAAT	CAAGGAGGAA	540
	TCCGGCATTG	CTATCTGCAC	CATGGTTTAC	CCTAGCGATG	AGAGCACCAA	ACTGAAGTCA	600
•	GCTGTCTTGA	CCCTGAAGGT	CATTCTGGGG	TTCTTCCTTC	CCTTCGTGGT	CATGGCTTGC	660
	TGCTATACCA	TCATCATTCA	CACCCTGATA	CAAGCCAAGA	AGTCTTCCAA	GCACAAAGCC	720
	CTAAAAGTGA	CCATCACTGT	CCTGACCGTC	TTTGTCTTGT	CTCAGTTTCC	CTACAACTGC	780
20	ATTTTGTTGG	TGCAGACCAT	TGACGCCTAT	GCCATGTTCA	TCTCCAACTG	TGCCGTTTCC	84.0
	ACCAACATTG	ACATCTGCTT	CCAGGTCACC	CAGACCATCG	CCTTCTTCCA	CAGTTGCCTG	900
	AACCCTGTTC	TCTATGTTTT	TGTGGGTGAG	AGATTCCGCC	GGGATCTCGT	GAAAACCCTG	960
	AAGAACTTGG	GTTGCATCAG	CCAGGCCCAG	TGGGTTTCAT	TTACAAGGAG	AGAGGGAAGC	1020
	TTGAAGCTGT	CGTCTATGTT	GCTGGAGACA	ACCTCAGGAG	CACTCTCCCT	CTGA	1074

- 25 (25) INFORMATION FOR SEQ ID NO:24:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 357 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
- 30 (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

	Met 1	Ala	Asp	Asp	Tyr 5	Gly	Ser	Glu	Ser	Thr 10	Ser	Ser	Met	Glu	Asp 15	Tyr
5	Val	Asn	Phe	Asn 20	Phe	Thr	Asp	Phe	Tyr 25	Cys	Glu	Lys	Asn	Asn 30	Val	Arg
	Gln	Phe	Ala 35	Ser	His	Phe	Leu	Pro 40	Pro	Leu	Tyr	Trp	Leu 45	Val	Phe	Ile
	Val	Gly 50	Ala	Leu	Gly	Asn	Ser 55	Leu	Val	Ile	Leu	Val 60	Tyr	Trp	Tyr	Cys
10	Thr 65	Arg	Val	Lys	Thr	Met 70	Thr	Asp	Met	Phe	Leu 75	Leu	Asn	Leu	Ala	Ile 80
	Ala	Asp	Leu	Leu	Phe 85	Leu	Val	Thr	Leu	Pro 90	Phe	Trp	Ala	Ile	Ala 95	Ala
15	Ala	Asp	Gln	Trp 100	Lys	Phe	Gln	Thr	Phe 105	Met	Cys	Lys	Val	Val	Asn	Ser
	Met	Tyr	Lys 115	Met	Asn	Phe	Tyr	Ser 120	Cys	Val	Leu	Leu	Ile 125	Met	Cys	Ile
	Ser	Val 130	Asp	Arg	Tyr	Ile	Ala 135	Ile	Ala	Gln	Ala	Met 140	Arg	Ala	His	Thr
20	Trp 145	Arg	Glu	Lys	Arg	Leu 150	Leu	Tyr	Ser	Lys	Met 155	Val	Cys	Phe	Thr	Ile 160
·	Trp	Val	Leu	Ala	Ala 165	Ala	Leu	Cys	Ile	Pro 170	Glu	Ile	Leu	Tyr	Ser 175	Gln
25	Ile	Lys	Glu	Glu 180	Ser	Gly	Ile	Ala	Ile 185	Cys	Thr	Met	Val	Tyr 190	Pro	Ser
	Asp	Glu	Ser 195	Thr	Lys	Leu	Lys	Ser 200	Ala	Val	Leu	Thr	Leu 205	Lys	Val	Ile
	Leu	Gly 210	P'ne	Phe	Leu	Pro	Phe 215	Val	Val	Met	Ala	Cys 220	Cys	Tyr	Thr.	Ile
30	Ile 225	Ile	His	Thr	Leu	Ile 230	Gln	Ala	Lys	Lys	Ser 235	Ser	Lys	His	Lys	Ala 240
	Leu	Lys	Val	Thr	Ile 245	Thr	Val	Leu	Thr	Val 250	Phe	Val	Leu	Ser	Gln 255	Phe
35	Pro	Tyr	Asn	Cys 260	Ile	Leu	Leu	Val	Gln 265	Thr	Ile	Asp	Ala	Tyr 270	Ala	Met
	Phe	Ile	Ser 275	Asn	Cys	Ala	Val	Ser 280	Thr	Asn	Ile	Asp	Ile 285	Cys	Phe	Gln

Val Thr Gln Thr Ile Ala Phe Phe His Ser Cys Leu Asn Pro Val Leu Tyr Val Phe Val Gly Glu Arg Phe Arg Arg Asp Leu Val Lys Thr Leu 310 5 - -Lys Asn Leu Gly Cys Ile Ser Gln Ala Gln Trp Val Ser Phe Thr Arg 330 325 Arg Glu Gly Ser Leu Lys Leu Ser Ser Met Leu Leu Glu Thr Thr Ser 345 Gly Ala Leu Ser Leu 10 355 (26) INFORMATION FOR SEQ ID NO:25: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1110 base pairs (B) TYPE: nucleic acid 15 (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (genomic) (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25: ATGGCCTCAT CGACCACTCG GGGCCCCAGG GTTTCTGACT TATTTTCTGG GCTGCCGCCG 60 20 GCGGTCACAA CTCCCGCCAA CCAGAGCGCA GAGGCCTCGG CGGGCAACGG GTCGGTGGCT GGCGCGGACG CTCCAGCCGT CACGCCCTTC CAGAGCCTGC AGCTGGTGCA TCAGCTGAAG 180 GGGCTGATCG TGCTGCTCTA CAGCGTCGTG GTGGTCGTGG GGCTGGTGGG CAACTGCCTG CTGGTGCTGG TGATCGCGCG GGTGCCGCGG CTGCACAACG TGACGAACTT CCTCATCGGC 300 AACCTGGCCT TGTCCGACGT GCTCATGTGC ACCGCCTGCG TGCCGCTCAC GCTGGCCTAT 25 GCCTTCGAGC CACGCGGCTG GGTGTTCGGC GGCGGCCTGT GCCACCTGGT CTTCTTCCTG CAGCCGGTCA CCGTCTATGT GTCGGTGTTC ACGCTCACCA CCATCGCAGT GGACCGCTAC GTCGTGCTGG TGCACCCGCT GAGGCGCGCA TCTCGCTGCG CCTCAGCCTA CGCTGTGCTG 600 GCCATCTGGG CGCTGTCCGC GGTGCTGCG CTGCCGCCCG CCGTGCACAC CTATCACGTG GAGCTCAAGC CGCACGACGT GCGCCTCTGC GAGGAGTTCT GGGGCTCCCA GGAGCGCCAG CGCCAGCTCT ACGCCTGGGG GCTGCTGCTG GTCACCTACC TGCTCCCTCT GCTGGTCATC 720

CTCCTGTCTT ACGTCCGGGT GTCAGTGAAG CTCCGCAACC GCGTGGTGCC GGGCTGCGTG

ACCCAGAGCC AGGCCGACTG GGACCGCGCT CGGCGCCGGC GCACCTTCTG CTTGCTGGTG

780

	GTGGTCGT	GG TG	GTGTTCG	C CGTC	TGCTGG	CTG	CCGC	TGC	ACGT	CTTC	AA C	CTGC	TGCG	G	900
	GACCTCGA	CC CC	CACGCCA	T CGAC	CCTTAC	GCC	TTTG	GGC	TGGT	'GCAG	CT G	CTCT	GCCA	.C	960
	TGGCTCGC	CA TG	AGTTCGG	C CTGC	TACAAC	ccc	TTCA	TCT	ACCC	CTGG	CT G	CACG	ACAG	C 1	020
	TTCCGCGA	GG AG	CTGCGCA	A ACTG	TTGGTC	GCT	TGGC	CCC	GCAA	GATA	.GC C	cccc	ATGG	C 1	080
5	CAGAATAT	GA CC	GTCAGCG	T GGTC	ATCTGA									1	110
	(27) INF	ORMAT:	ION FOR	SEQ I	D NO:2	6 :			•						
10	·	(A) (B) (C) (D)	UENCE C LENGTH TYPE: STRAND TOPOLOG	: 369 amino EDNESS GY: no	amino acid : t rele	acid vant	s								• •
			QUENCE I	·	-		ID 1	NO : 2	6 :						
			Ser Ser			_				Ser	Asp	Len	Phe	Ser	
15	1		,	5	J	1		10			p	Deu	15	Ser	-
	Gly	Leu I	Pro Pro 20	Ala V	al Thr	Thr	Pro 25	Ala	Asn	Gln	Ser	Ala 30	Glu	Ala	4,
	Ser		Gly Asn B5	Gly S	er Val	Ala 40	Gly	Ala	Asp	Ala	Pro 45	Ala	Val	Thr	ž
20	Pro	Phe G	Sln Ser	Leu G	ln Leu 55	Val	His	Gln	Leu	Lys 60	Gly	Leu	Ile	Val	v
	Leu 65	Leu I	Tyr Ser	Val Va		Val	Val	Gly	Leu 75	Val	Gly	Asn	Cys	Leu 80	
25	Leu	Val I	Leu Val	Ile Al	la Arg	Val	Pro	Arg 90	Leu	His	Asn	Val	Thr 95	Asn	
	Phe	Leu I	le Gly	Asn Le	eu Ala	Leu	Ser 105	Asp	Val	Leu	Met	Cys 110	Thr	Ala	
	Cys		Pro Leu .15	Thr Le	eu Ala	Tyr 120	Ala	Phe	Glu	Pro	Arg 125	Gly	Trp	Val	
30	Phe	Gly G 130	sly Gly	Leu Cy	s His	Leu	Val	Phe	Phe	Leu 140	Gln	Pro	Val	Thr	
	Val 145	Tyr V	al Ser	Val Ph		Leu	Thr	Thr	Ile 155	Ala	Val	Asp	Arg	Tyr 160	
35	Val	Val L	eu Val	His Pr 165	o Leu	Arg	Arg	Ala 170	Ser	Arg	Cys	Ala	Ser 175	Ala	

		Tyŗ	Ala	Val	Leu 180	Ala	Ile	Trp	Ala	Leu 185	Ser	Ala	Val	Leu	Ala 190	Leu	Pro	
		Pro	Ala	Val 195	His	Thr	Tyr	His	Val 200	Glu	Leu	Lys	Pro	His 205	Asp	Val	Arg	
5	•	Leu	Cys 210	Glu	Glu	Phe.	Trp.	Gly 215	Ser	Gln	Glu	Arg	Gln 220	Arg	Gln	Leu	Tyr	
		Ala 225	Trp	Gly	Leu	Leu	Leu 230	Val	Thr	Tyr	Leu	Leu 235	Pro	Leu	Leu	Val	Ile 240	
10		Leu	Leu	Ser	Tyr	Val 245	Arg	Val	Ser	Val	Lys 250	Leu	Arg	Asn	Arg	Val 255	Val	
		Pro	Gly	Cys	Val 260	Thr	Gln	Ser	Gln	Ala 265	Asp	Trp	Asp	Arg	Ala 270	Arg	Arg	
		Arg	Arg	Thr 275	Phe	Cys	Leu	Leu	Val 280	Val	Val	Val	Val	Val 285	Phe	Ala	Val	
15	•	Cys	Trp 290	Leu	Pro	Leu	His	Val 295	Phe	Asn	Leu	Leu	Arg 300	Asp	Leu	Asp	Pro	
		His 305	Ala	Ile	Asp	Pro	Tyr 310	Ala	Phe	Gly	Leu	Val 315	Gln	Leu	Leu	Cys	His 320	٠
20		Trp	Leu	Ala	Met	Ser 325	Ser	Ala	Cys	Tyr	Asn 330	Pro	Phe	Ile	Tyr	Ala 335	Trp	
		Leu	His	Asp	Ser 340	Phe	Arg	Glu	Glu	Leu 345	Arg	Lys	Leu	Leu	Val 350	Ala	Trp	
•		Pro	Arg	Lys 355	Ile	Ala	Pro	His	Gly 360	Gln	Asn	Met	Thr	Val 365	Ser	Val	Val	
25		Ile																
	(28)	INF	ORMA!	rion	FOR	SEQ	ID 1	NO:27	7:									
30		(i)	(A) (B) (C)	LEI TYI	CE CH NGTH PE: 1 RANDI POLO	: 108 nucle EDNES	33 ba eic a SS: s	ase p acid sing!	pairs	6								
	•	(i:	i) M	OLECT	ULE :	TYPE	: DN	A (ge	enom:	ic)								
		(x:	i) SI	EQUEI	NCE I	DESCI	RIPT	ION:	SEQ	ID I	NO : 2	7:						
35	ATGG	ACCC	AG A	AGAA	ACTT(C AG	TTTA!	TTTG	GAT'	TATT?	ACT I	ATGC:	racg/	AG C	CCAA	ACTC:	Γ	60

GACATCAGGG AGACCCACTC CCATGTTCCT TACACCTCTG TCTTCCTTCC AGTCTTTTAC 120

WO 00/22129 PCT/US99/23938

23

-	ACAGCTGTGT	TCCTGACTGG	AGTGCTGGGG	AACCTTGTTC	TCATGGGAGC	GTTGCATTTC	180
	AAACCCGGCA	GCCGAAGACT	GATCGACATC	TTTATCATCA	ATCTGGCTGC	CTCTGACTTC	240
	ATTTTTCTTG	TCACATTGCC	TCTCTGGGTG	GATAAAGAAG	CATCTCTAGG	ACTGTGGAGG	300
	ACGGGCTCCT	TCCTGTGCAA	AGGGAGCTCC	TACATGATCT	CCGTCAATAT	GCACTGCAGT	360
5	GTCCTCCTGC	TCACTTGCAT	GAGTGTTGAC	CGCTACCTGG	CCATTGTGTG	GCCAGTCGTA	420
	TCCAGGAAAT	TCAGAAGGAC	AGACTGTGCA	TATGTAGTCT	GTGCCAGCAT	CTGGTTTATC	480
	TCCTGCCTGC	TGGGGTTGCC	TACTCTTCTG	TCCAGGGAGC	TCACGCTGAT	TGATGATAAG	540
	CCATACTGTG	CAGAGAAAAA	GGCAACTCCA	ATTAAACTCA	TATGGTCCCT	GGTGGCCTTA	600
	ATTTTCACCT	TTTTTGTCCC	TTTGTTGAGC	ATTGTGACCT	GCTACTGTTG	CATTGCAAGG	660
10	AAGCTGTGTG	CCCATTACCA	GCAATCAGGA	AAGCACAACA	AAAAGCTGAA	GAAATCTATA	720
	AAGATCATCT	TTATTGTCGT	GGCAGCCTTT	CTTGTCTCCT	GGCTGCCCTT	CAATACTTTC	780
	AAGTTCCTGG	CCATTGTCTC	TGGGTTGCGG	CAAGAACACT	ATTTACCCTC	AGCTATTCTT	840
	CAGCTTGGTA	TGGAGGTGAG	TGGACCCTTG	GCATTTGCCA	ACAGCTGTGT	CAACCCTTTC	900
	ATTTACTATA	TCTTCGACAG	CTACATCCGC	CGGGCCATTG	TCCACTGCTT	GTGCCCTTGC	960
15	CTGAAAAACT	ATGACTTTGG	GAGTAGCACT	GAGACATCAG	ATAGTCACCT	CACTAAGGCT	1020
	CTCTCCACCT	TCATTCATGC	AGAAGATTTT	GCCAGGAGGA	GGAAGAGGTC	TGTGTCACTC	1080
	TAA						1083
	(29) INFORM	MATION FOR S	EQ ID NO:28):			
20		EQUENCE CHA					

- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Met Asp Pro Glu Glu Thr Ser Val Tyr Leu Asp Tyr Tyr Tyr Ala Thr 1 5 10 15

Ser Pro Asn Ser Asp Ile Arg Glu Thr His Ser His Val Pro Tyr Thr 20 25 30

30 Ser Val Phe Leu Pro Val Phe Tyr Thr Ala Val Phe Leu Thr Gly Val 35 40 45

		Leu	Gly 50	Asn	Leu	Val	Leu	Met 55	Gly	Ala	Leu	His	Phe	Lys	Pro	Gly	Ser
		Arg 65	Arg	Leu	Ile	Asp	Ile 70	Phe	Ile	Ile	Asn	Leu 75	Ala	Ala	Ser	Asp	Phe 80
5		Ile	Phe	Leu	Val	Thr 85	Leu	Pro	Leu	Trp	Val 90	Asp	Lys	Glu	Ala	Ser 95	Leu
		Gly	Leu	Trp	Arg 100	Thr	Gly	Ser	Phe	Leu 105	Cys	Lys	Gly	Ser	Ser 110	Tyr	Met
10		Ile	Ser	Val 115	Asn	Met	His	Cys	Ser 120	Val	Leu	Leu	Leu	Thr 125	Cys	Met	Ser
		Val	Asp 130	Arg	Tyr	Leu	Ala	Ile 135	Val	Trp	Pro	Val	Val 140	Ser	Arg	Lys	Phe
		Arg 145	Arg	Thr	Asp	Cys	Ala 150	Tyr	Val	Val	Cys	Ala 155	Ser	Ile	Trp	Phe	Ile 160
15		Ser	Cys	Leu	Leu	Gly 165	Leu	Pro	Thr	Leu	Leu 170	Ser	Arg	Glu	Leu	Thr 175	Leu
		Ile	Asp	Asp	Lys 180	Pro	Tyr	Cys	Ala	Glu 185	Lys	Lys	Ala	Thr	Pro 190	Ile	Lys
20		Leu	Ile	Trp 195	Ser	Leu	Val	Ala	Leu 200	Ile	Phe	Thr	Phe	Phe 205	Val	Pro	Leu
		Leu	Ser 210	Ile	Val	Thr	Cys	Tyr 215	Cys	Cys	Ile _.	Ala	Arg 220	Lys	Leu	Cys	Ala
		His 225	Tyr	Gln	Gln	Ser	Gly 230	Lys	His	Asn	Lys	Lys 235	Leu	Lys	Lys	Ser	Ile 240
25		Lys	Ile	Ile	Phe	Ile 245	Val	Val	Ala	Ala	Phe 250	Leu	Val	Ser	Trp	Leu 255	Pro
		Phe	Asn	Thr	Phe 260	Lys	Phe	Leu	Ala	Ile 265	Val	Ser	Gly	Leu	Arg 270	Gln	Glu
30		His	Tyr	Leu 275	Pro	Ser	Ala	Ile	Leu 280	Gln	Leu	Gly	Met	Glu 285	Val	Ser	Gly
		Pro	Leu 290	Ala	Phe	Ala	Asn	Ser 295	Cys	Val	Asn	Pro	Phe 300	Ile	Tyr	Tyr	Ile
	• .	Phe 305	Asp	Ser	Tyr	Ile	Arg 310	Arg	Ala	Ile	Val	His 315	Cys	Leu	Cys	Pro	Cys 320
35		Leu	Lys	Asn	Tyr	Asp 325	Phe	Gly	Ser	Ser	Thr 330	Glu	Thr	Ser	Asp	Ser 335	His
		Leu	Thr	Lys	Ala	Leu	Ser	Thr	Phe	Ile	His	Ala	Glu	Asp	Phe	Ala	Arg

PCT/US99/23938

	•		
,	345	350	
	Arg Arg Lys Arg Ser Val Ser Leu 355 360		
	(30) INFORMATION FOR SEQ ID NO:29:	·	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 		·
10	(ii) MOLECULE TYPE: DNA (genomic)	·	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:	•	
	CTAGAATTCT GACTCCAGCC AAAGCATGAA T		31
	(31) INFORMATION FOR SEQ ID NO:30:	•	
15	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs		t
	(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear		
	(ii) MOLECULE TYPE: DNA (genomic)		Ą
20	(x1) SEQUENCE DESCRIPTION: SEQ ID NO:30:		
	GCTGGATCCT AAACAGTCTG CGCTCGGCCT		30
-	(32) INFORMATION FOR SEQ ID NO:31:		
25	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1020 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 		
	(ii) MOLECULE TYPE: DNA (genomic)		
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:		
30	ATGAATGGCC TTGAAGTGGC TCCCCCAGGT CTGATCACCA ACTTCTCCCT	GGCCACGGCA	60
	GAGCAATGTG GCCAGGAGAC GCCACTGGAG AACATGCTGT TCGCCTCCTT	CTACCTTCTG	120
	GATTTTATCC TGGCTTTAGT TGGCAATACC CTGGCTCTGT GGCTTTTCAT	CCGAGACCAC	180
	AAGTCCGGGA CCCCGGCCAA CGTGTTCCTG ATGCATCTGG CCGTGGCCGA	CTTGTCGTGC	240
	GTGCTGGTCC TGCCCACCCG CCTGGTCTAC CACTTCTCTG GGAACCACTG	GCCATTTGGG	300

	GAAATCGCAT	GCCGTCTCAC	CGGCTTCCTC	TTCTACCTCA	ACATGTACGC	CAGCATCTAC	360
	TTCCTCACCT	GCATCAGCGC	CGACCGTTTC	CTGGCCATTG	TGCACCCGGT	CAAGTCCCTC	420
	AAGCTCCGCA	GGCCCCTCTA	CGCACACCTG	GCCTGTGCCT	TCCTGTGGGT	GGTGGTGGCT	480
	GTGGCCATGG	CCCCGCTGCT	GGTGAGCCCA	CAGACCGTGC	AGACCAACCA	CACGGTGGTC	540
5	TGCCTGCAGC	TGTACCGGGA	GAAGGCCTCC	CACCATGCCC	TGGTGTCCCT	GGCAGTGGCC	600
	TTCACCTTCC	CGTTCATCAC	CACGGTCACC	TGCTACCTGC	TGATCATCCG	CAGCCTGCGG	660
	CAGGGCCTGC	GTGTGGAGAA	GCGCCTCAAG	ACCAAGGCAG	TGCGCATGAT	CGCCATAGTG	720
	CTGGCCATCT	TCCTGGTCTG	CTTCGTGCCC	TACCACGTCA	ACCGCTCCGT	CTACGTGCTG	780
	CACTACCGCA	GCCATGGGGC	CTCCTGCGCC	ACCCAGCGCA	TCCTGGCCCT	GGCAAACCGC	840
10	ATCACCTCCT	GCCTCACCAG	CCTCAACGGG	GCACTCGACC	CCATCATGTA	TTTCTTCGTG	900
	GCTGAGAAGT	TCCGCCACGC	CCTGTGCAAC	TTGCTCTGTG	GCAAAAGGCT	CAAGGGCCCG	960
e.	CCCCCCAGCT	TCGAAGGGAA	AACCAACGAG	AGCTCGCTGA	GTGCCAAGTC	AGAGCTGTGA	1020
	(33) INFORM	MATION FOR S	SEQ ID NO:32	2:			

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 339 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein
- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Met Asn Gly Leu Glu Val Ala Pro Pro Gly Leu Ile Thr Asn Phe Ser 1 5 10 15

Leu Ala Thr Ala Glu Gln Cys Gly Gln Glu Thr Pro Leu Glu Asn Met 20 25 30

Leu Phe Ala Ser Phe Tyr Leu Leu Asp Phe Ile Leu Ala Leu Val Gly 35 40 45

Asn Thr Leu Ala Leu Trp Leu Phe Ile Arg Asp His Lys Ser Gly Thr 50 55 60

Pro Ala Asn Val Phe Leu Met His Leu Ala Val Ala Asp Leu Ser Cys 65 70 75 80

Val Leu Val Leu Pro Thr Arg Leu Val Tyr His Phe Ser Gly Asn His 85 90 95

Trp Pro Phe Gly Glu Ile Ala Cys Arg Leu Thr Gly Phe Leu Phe Tyr

	•				100		•			105				•	110		
		Leu	Asn	Met 115	Tyr	Ala	Ser	Ile	Tyr 120	Phe	Leu	Thr	Cys	Ile 125	Ser	Ala	Asp
5		Arg	Phe 130	Leu	Ala	Ile	Val	His 135	Pro	Val	Lys	Ser	Leu 140	Lys	Leu	Arg	Arg
		Pro 145	Leu	Tyr	Ala	His	Leu 150	Ala	Cys	Ala	Phe	Leu 155	Trp	Val	Val	Val	Ala 160
		Val	Ala	Met	Ala	Pro 165	Leu	Leu	Val	Ser	Pro 170	Gln	Thr	Val	Gln	Thr 175	Asn
10		His	Thr	Val	Val 180	Cys	Leu	Gln	Leu	Tyr 185	Arg	Glu	Lys	Ala	Ser 190	His	His
		Ala	Leu	Val 195	Ser	Leu	Ala	Val	Ala 200	Phe	Thr	Phe	Pro	Phe 205	Ile	Thr	Thr
15		Val	Thr 210	Cys	Tyr	Leu	Leu	Ile 215	Ile	Arg	Ser	Leu	Arg 220	Gln	Gly	Leu	Arg
		Val 225	Glu	Lys	Arg	Leu	Lys 230	Thr	Lys	Ala	Val	Arg 235	Met	Ile	Ala	Ile	Val 240
		Leu	Ala	Ile	Phe	Leu 245	Val	Cys	Phe	Val	Pro 250	Tyr	His	Val	Asn	Arg 255	Ser
20		Val	Tyr	Val	Leu 260	His	Tyr	Arg	Ser	His 265	Gly	Ala	Ser	Cys	Ala 270	Thr	Gln
		Arg	Ile	Leu 275	Ala	Leu	Ala	Asn	Arg 280	Ile	Thr	Ser	Cys	Leu 285	Thr	Ser	Leu
25 -		Asn	Gly 290	Ala	Leu	Asp	Pro	Ile 295	Met	Tyr	Phe	Phe	Val 300	Ala	Glu	Lys	Phe
		Arg 305	His	Ala	Leu	Cys	Asn 310	Leu	Leu	Cys	Gly	Lys 315	Arg	Leu	Lys	Gly	Pro 320
		Pro	Pro	Ser	Phe	Glu 325	Gly	Lys	Thr	Asn	Glu 330	Ser	Ser	Leu	Ser	Ala 335	Lys
30	(34)		Glu RMAT		FOR	SEQ	ID N	10:33	i:								
35		(i)	(B)	UENC LEN TYP STR	GTH: E: n	29 ucle	base	pai cid	.rs								
			(D)	TOP	OLOG	Y: 1	inea	ır									

(ii) MOLECULE TYPE: DNA (genomic)

	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:	
	ATAAGATGAT CACCCTGAAC AATCAAGAT	29
	(35) INFORMATION FOR SEQ ID NO:34:	
. 5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	•
	(ii) MOLECULE TYPE: DNA (genomic)	
10	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:	
	TCCGAATTCA TAACATTTCA CTGTTTATAT TGC	33
	(36) INFORMATION FOR SEQ ID NO:35:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 996 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	-
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:	
20	ATGATCACCC TGAACAATCA AGATCAACCT GTCACTTTTA ACAGCTCACA TCCAGATGAA	60
	TACAAAATTG CAGCCCTTGT CTTCTATAGC TGTATCTTCA TAATTGGATT ATTTGTTAAC	120
	ATCACTGCAT TATGGGTTTT CAGTTGTACC ACCAAGAAGA GAACCACGGT AACCATCTAT	180
	ATGATGAATG TGGCATTAGT GGACTTGATA TTTATAATGA CTTTACCCTT TCGAATGTTT	240
	TATTATGCAA AAGATGCATG GCCATTTGGA GAGTACTTCT GCCAGATTAT TGGAGCTCTC	300
25	ACAGTGTTTT ACCCAAGCAT TGCTTTATGG CTTCTTGCCT TTATTAGTGC TGACAGATAC	360
	ATGGCCATTG TACAGCCGAA GTACGCCAAA GAACTTAAAA ACACGTGCAA AGCCGTGCTG	420
	GCGTGTGTGG GAGTCTGGAT AATGACCCTG ACCACGACCA CCCCTCTGCT ACTGCTCTAT	480
	AAAGACCCAG ATAAAGACTC CACTCCCGCC ACCTGCCTCA AGATTTCTGA CATCATCTAT	540
	CTAAAAGCTG TGAACGTGCT GAACCTCACT CGACTGACAT TTTTTTTCTT GATTCCTTTG	600
30	TTCATCATGA TTGGGTGCTA CTTGGTCATT ATTCATAATC TCCTTCACGG CAGGACGTCT	660

AAGCTGAAAC CCAAAGTCAA GGAGAAGTCC ATAAGGATCA TCATCACGCT GCTGGTGCAG 720

	GTGCT	CG?	rct (GCTT'	TATG	cc c	TTCC.	ACAT	C TG	TTTC	GCTT	TCC	TGAT	GCT (GGGA	ACGG	GG	780
	GAGAAC	CAC	STT 2	ACAA:	TCCC'	rg g	GGAG	CCTT	r ac	CACC'	TTCC	TCA	TGAA	CCT (CAGC	ACGT	GT	840
	CTGGAT	rgi	rga :	r tct (CTAC'	ra c	ATCG:	rttc <i>i</i>	A AA	ACAA'	TTTC	AGG	CTCG	AGT (CATT	AG T G'	rc	900
	ATGCTA	\T#	ACC (TAA!	TTAC	CT T	CGAA	CCT	G CG	CAGA	AAAA	GTT'	rccg	ATC :	rggt <i>i</i>	AGTC:	ΓA	960
5	AGGTCA	CI	AA.	CAAT	[ATA]	AA C	AGTG2	AAATO	3 T T2	ATGA								996
	(37) I	NF	ORMA	TION	I FOI	R SE	Q ID	NO:3	6:									
0			(A (E (C	A) LE B) TY C) SI D) TO	ENGTH PE: RANI POLC	H: 33 amir EDNE GY:	31 am no ac ESS: not	rele	ació									
		(i	i) M	OLEC	ULE	TYPE	: pr	otei	n.									
		(x	i) S	EQUE	NCE	DESC	RIPT	: NOI	SEC	ID	NO: 3	6:						
5	M 1	et	Ile	Thr	Leu	Asn 5	Asn	Gln	Asp	Gln	Pro 10	Val	Thr	Phe	Asn	Ser 15	Ser	
	H	is	Pro	Asp	Glu 20	Tyr	Lys	Ile	Ala	Ala 25	Leu	Val	Phe	Tyr	Ser 30	Cys	Ile	
	P)	he	Ile	Ile 35	Gly	Leu	Phe	Val	Asn 40	Ile	Thr	Ala	Leu	Trp 45	Val	Phe	Ser	
0	C	ys	Thr 50	Thr	Lys	Lys	Arg	Thr 55	Thr	Val	Thr	Ile	Tyr 60	Met	Met	Asn	Val	er - Zu
	A]	la 5	Leu	Val	Asp	Leu	Ile 70	Phe	Ile	Met	Thr	Leu 75	Pro	Phe	Arg	Met	Phe 80	
5	Τ <u>)</u>	yr	Tyr	Ala	Lys	Asp 85	Ala	Trp	Pro	Phe	Gly 90	Glu	Tyr	Phe	Cys	Gln 95	Ile	
	13	le	Gly	Ala	Leu 100	Thr	Val	Phe	Tyr	Pro 105	Ser	Ile	Ala	Leu	Trp 110	Leu	Leu	
	· Al	La	Phe	Ile 115	Ser	Ala	Asp	Arg	Tyr 120	Met	Ala	Ile	Val	Gln 125	Pro	Lys	Tyr	
)	Al	.a	Lys 130	-Glu	Leu	Lys	Asn	Thr 135	Cys	Lys	Ala	Val	Leu 140	Ala	Cys	Val	Gly	
	Va 14	1	Trp	Ile	Met	Thr	Leu 150	Thr	Thr	Thr	Thr	Pro 155	Leu	Leu	Leu	Leu	Tyr 160	
5	Ly	s	Asp	Pro	Asp	Lys 165	Asp	Ser	Thr	Pro	Ala 170	Thr	Cys	Leu	Lys	Ile 175	Ser	

	Asp	Ile	Ile	Tyr 180	Leu	Lys	Ala	Val	Asn 185	Val	Leu	Asn	Leu	Thr 190	Arg	Leu
	Thr	Phe	Phe 195	Phe	Leu	Ile	Pro	Leu 200	Phe	Ile	Met	Ile	Gly 205	Cys	Tyr	Leu
5	Val	Ile 210	Ile	His	Asn	Leu	Leu 215	His	Gly	Arg	Thr	Ser 220	Lys	Leu	Lys	Pro
	Lys 225	Val	Lys	Glu	Lys	Ser 230	Ile	Arg	Ile	Ile	Ile 235	Thr	Leu	Leu	Val	Gln 240
10	Val	Leu	Val	Cys	Phe 245	Met	Pro	Phe	His	Ile 250	Cys	Phe	Ala	Phe	Leu 255	Met
	Leu	Gly	Thr	GÌy 260	Glu	Asn	Ser	Tyr	Asn 265	Pro	Trp	Gly	Ala	Phe 270	Thr	Thr
	Phe	Leu	Met 275	Asn	Leu	Ser	Thr	Cys 280	Leu	Asp	Val	Ile	Leu 285	Tyr	Tyr	Ile
15	Val	Ser 290	Lys	Gln	Phe	Gln	Ala 295	Arg	Val	Ile.	Ser	Val 300	Met	Leu	Tyr	Arg
	Asn 305	Tyr	Leu	Arg	Ser	Leu 310	Arg	Arg	Lys	Ser	Phe 315	Arg	Ser	Gly	Ser	Leu 320
20	Arg	Ser	Leu	Ser	Asn 325	Ile	Asn	Ser	Glu	Met 330	Leu					
	(38) INF	ORMA!	rion	FOR	SEQ	ID 1	/O:3	7:						*		
25	(i	(B)) LEI) TYI) STI	CE CI NGTH PE: 1 RANDI POLO	: 28 nucle EDNES	base eic a SS: a	e pa: acid sing:	irs						· ;		
	(i	i) M	OLEC	ULE :	TYPE	: DN	A (g	enom:	ic)							
	(x	i) SI	EQUEI	NCE 1	DESCI	RIPT	ION:	SEQ	ID I	NO:37	7:					
	CCAAGCTT	CC A	GGCC	TGGG	G TG	rgc t (G G									28
30	(39) INF	ORMA'	TION	FOR	SEQ	ID 1	NO:3	B:								
35	(i	(B (C) LE	CE CI NGTH PE: 1 RANDI POLO	: 29 nucle EDNE:	base eic a	e pa: acid sing	irs								
	(i	i) M	OLEC	ULE '	TYPE	: DN	A (g	enom:	ic)			•				•
	(*	i) S	EOUE	NCE :	DESC	RIPT	TON:	SEO	TD 1	NO : 31	я.					

10

960

1077

ATGGATCCTG ACCTTCGGCC CCTGGCAGA 29 (40) INFORMATION FOR SEQ ID NO:39: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1077 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (genomic) (xi) SEQUENCE DESCRIPTION: SEQ ID NO:39: ATGCCCTCTG TGTCTCCAGC GGGGCCCTCG GCCGGGGCAG TCCCCAATGC CACCGCAGTG 60 ACAACAGTGC GGACCAATGC CAGCGGGCTG GAGGTGCCCC TGTTCCACCT GTTTGCCCGG 120 CTGGACGAGG AGCTGCATGG CACCTTCCCA GGCCTGTGCG TGGCGCTGAT GGCGGTGCAC 180 GGAGCCATCT TCCTGGCAGG GCTGGTGCTC AACGGGCTGG CGCTGTACGT CTTCTGCTGC 240 CGCACCCGGG CCAAGACACC CTCAGTCATC TACACCATCA ACCTGGTGGT GACCGATCTA 300 CTGGTAGGGC TGTCCCTGCC CACGCGCTTC GCTGTGTACT ACGGCGCCAG GGGCTGCCTG 360 CGCTGTGCCT TCCCGCACGT CCTCGGTTAC TTCCTCAACA TGCACTGCTC CATCCTCTTC 420 CTCACCTGCA TCTGCGTGGA CCGCTACCTG GCCATCGTGC GGCCCGAAGG CTCCCGCCGC 480 TGCCGCCAGC CTGCCTGTGC CAGGGCCGTG TGCGCCTTCG TGTGGCTGGC CGCCGGTGCC 540 GTCACCCTGT CGGTGCTGGG CGTGACAGGC AGCCGGCCCT GCTGCCGTGT CTTTGCGCTG 600 ACTGTCCTGG AGTTCCTGCT GCCCCTGCTG GTCATCAGCG TGTTTACCGG CCGCATCATG 660 TGTGCACTGT CGCGGCCGGG TCTGCTCCAC CAGGGTCGCC AGCGCCGCGT GCGGGCCATG 720 CAGCTCCTGC TCACGGTGCT CATCATCTTT CTCGTCTGCT TCACGCCCTT CCACGCCCGC 780 840 GTGGCCGTGA CCCTCAGCAG CCTCAACAGC TGCATGGACC CCATCGTCTA CTGCTTCGTC 900

ACCAGTGGCT TCCAGGCCAC CGTCCGAGGC CTCTTCGGCC AGCACGGAGA GCGTGAGCCC

CTCAGTGCCG GCCCTCACGC CCTCACCCAG GCCCTGGCTA ATGGGCCCGA GGCTTAG

AGCAGCGGTG ACGTGGTCAG CATGCACAGG AGCTCCAAGG GCTCAGGCCG TCATCACATC 1020

(41) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 358 amino acids

30

25

· .	. :	(C) TY:) ST:) TO:	RAND	EDNE	SS:			٠.		, •					
	(i:	i) M	OLEC	ULE '	TYPE	: pr	otei	n.								
5	, (x :	i) S	EQUEI	NCE 1	DESC	RIPT:	ION:	SEQ	ID I	NO : 4	D:					
	Met 1	Pro	Ser	Val	Ser 5	Pro	Ala	Gly	Pro	Ser 10	Ala	Gly	Ala	Val	Pro 15	Asn
	Ala	Thr	Ala	Val 20	Thr	Thr	Val	Arg	Thr 25	Asn	Ala	Ser	Gly	Leu 30	Glu	Val
10	Pro	Leu	Phe 35	His	Leu	Phe		Arg 40	Leu	Asp	Glu	Glu	Leu 45	His	Gly	Thr
	Phe	Pro 50	Gly	Leu	Cys	Val	Ala 55	Leu	Met	Ala	Val	His 60	Gly	Ala	Ile	Phe
15	Leu 65	Ala	Gly	Leu	Val	Leu 70	Asn	Gly	Leu	Ala	Leu 75	Tyr	Val	Phe	Cys	Cys 80
	Arg	Thr	Arg	Ala	Lys 85	Thr	Pro	Ser	Val	Ile 90	Tyr	Thr	Ile	Asn	Leu 95	Val
	Val	Thr	Asp	Leu 100	Leu	Val	Gly	Leu	Ser 105	Leu	Pro	Thr	Arg	Phe 110	Ala	Val
20	Tyr	Tyr	Gly 115	Ala	Arg	Gly	Cys	Leu 120	Arg	Cys	Ala	Phe	Pro 125	His	Val	Leu
	Gly	Tyr 130	Phe	Leu	Asn	Met	His 135	Cys	Ser	Ile	Leu	Phe 140	Leu	Thr	Cys	Ile
25	Cys 145	Val	Asp	Arg	Tyr	Leu 150	Ala	Ile	Val	Arg	Pro 155	Glu	Ala	Pro	Ala	Ala 160
	Cys	Arg	Gln	Pro	Ala 165	Cys	Ala	Arg	Ala	Val 170	Cys	Ala	Phe	Val	Trp 175	Leu
	Ala	Ala	Gly	Ala 180	Val	Thr	Leu	Ser	Val 185	Leu	Gly	Val	Thr	Gly 190	Ser	Arg
30	Pro	Cys	Cys 195	Arg	Val	Phe	Ala	Leu 200	Thr	Val	Leu	Glu	Phe 205	Leu	Leu	Pro
	Leu	Leu 210	Val	Ile	Ser	Val	Phe 215	Thr	Gly	Arg	Ile	Met 220	Cys	Ala	Leu	Ser
35	Arg 225	Pro	Gly	Leu	Leu	His 230	Gln	Gly	Arg	Gln	Arg 235	Arg	Val	Arg	Ala	Met 240
	Gln	Leu	Leu	Leu	Thr	Val	Leu	Ile	Ile	Phe	Leu	Val	Cys	Phe	Thr	Pro

בו בשמנומים אותר מואים בשוב ו -

PCT/US99/23938

		Phe	His	Ala	Arg 260	Gln	Val	Ala	Val	Ala 265	Leu	Trp	Pro	Asp	Met 270	Pro	His	
		His	Thr	Ser 275	Leu	Val	Val	Tyr	His 280	Val	Ala	Val	Thr	Leu 285	Ser	Ser	Leu	
5		Asn	Ser 290	Cys	Met	Asp	Pro	Ile 295	Val	Tyr	Cys	Phe	Val 300	Thr	Ser	Gly	Phe	
		Gln 305	Ala	Thr	Val	Arg	Gly 310	Leu	Phe	Gly	Gln	His 315	Gly	Glu	Arg	Glu	Pro 320	
10		Ser	Ser	Gly	Asp	Val 325	Val	Ser	Met	His	Arg 330	Ser	Ser	Lys	Gly	Ser 335	Gly	
		Arg	His	His	Ile 340	Leu	Ser	Ala	Gly	Pro 345	His	Ala	Leu	Thr	Gln 350	Ala	Leu	
		Ala	Asn	Gly 355	Pro	Glu	Ala		•									
15	(42)	INFO	RMAT	NOI	FOR	SEQ	ID 1	10:43	:	• •	·				•			
20		(i)	(A) (B) (C)	LEN TYI STI	CE CH NGTH: PE: n RANDE	30 nucle	base ic a SS: s	e pai acid singl	rs									
		(ii	.) MC	LECU	JLE I	YPE:	DNZ	(ge	nomi	.c)								ı
		(xi	.) SE	EQUEN	ICE I	ESCF	RIPTI	ON:	SEQ	ID N	IO:41	. :						
	GAGAZ	ATTCA	C TO	CTG	GCTC	: AAG	ATG	AACT										30
	(43)	INFO	RMAI	CION	FOR	SEQ	ID 1	10:42	2 :									
25		(i)	(A) (B) (C)	LEN TYP STR	CE CH IGTH: PE: n RANDE	30 ucle DNES	base ic a SS: s	e pai acid singl	.rs					·				
30		(ii	.) MC	LECU	JLE I	YPE:	DNA	(ge	enomi	.c)								
		(xi) SI	EQUEN	ICE I	ESCF	(IPT	ON:	SEQ	ID N	10:42	::						
	CGGGZ	ATCCC	C GI	TAACT	rgago	CAC	TTC	AGAT										30
	(44)	INFO	RMAT	CION	FOR	SEQ	ID 1	10:43	3:									
35		(i)	(A) (B)	LEN	CE CH IGTH: PE: I	105 ucle	0 ba	ase p acid	airs	:								

(D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRITTION: SEQ ID NO:43:

	ATGAACTCCA	CCTTGGATGG	TAATCAGAGC	AGCCACCCTT	TTTGCCTCTT	GGCATTTGGC	60
5	TATTTGGAAA	CTGTCAATTT	TTGCCTTTTG	GAAGTATTGA	TTATTGTCTT	TCTAACTGTA	120
	TTGATTATTT	CTGGCAACAT	CATTGTGATT	TTTGTATTTC	ACTGTGCACC	TTTGTTGAAC	180
	CATCACACTA	CAAGTTATTT	TATCCAGACT	ATGGCATATG	CTGACCTTTT	TGTTGGGGTG	240
	AGCTGCGTGG	TCCCTTCTTT	ATCACTCCTC	CATCACCCCC	TTCCAGTAGA	GGAGTCCTTG	300
	ACTTGCCAGA	TATTTGGTTT	TGTAGTATCA	GTTCTGAAGA	GCGTCTCCAT	GGCTTCTCTG	360
10	GCCTGTATCA	GCATTGATAG	ATACATTGCC	ATTACTAAAC	CTTTAACCTA	TAATACTCTG	420
	GTTACACCCT	GGAGACTACG	CCTGTGTATT	TTCCTGATTT	GGCTATACTC	GACCCTGGTC	480
	TTCCTGCCTT	CCTTTTTCCA	CTGGGGCAAA	CCTGGATATC	ATGGAGATGT	GTTTCAGTGG	540
	TGTGCGGAGT	CCTGGCACAC	CGACTCCTAC	TTCACCCTGT	TCATCGTGAT	GATGTTATAT	600
	GCCCCAGCAG	CCCTTATTGT	CTGCTTCACC	TATTTCAACA	TCTTCCGCAT	CTGCCAACAG	660
15	CACACAAAGG	ATATCAGCGA	AAGGCAAGCC	CGCTTCAGCA	GCCAGAGTGG	GGAGACTGGG	720
	GAAGTGCAGG	CCTGTCCTGA	TAAGCGCTAT	GCCATGGTCC	TGTTTCGAAT	CACTAGTGTA	780
	TTTTACATCC	TCTGGTTGCC	ATATATCATC	TACTTCTTGT	TGGAAAGCTC	CACTGGCCAC	840
	AGCAACCGCT	TCGCATCCTT	CTTGACCACC	TGGCTTGCTA	TTAGTAACAG	TTTCTGCAAC	900
	TGTGTAATTT	ATAGTCTCTC	CAACAGTGTA	TTCCAAAGAG	GACTAAAGCG	CCTCTCAGGG	960
20	GCTATGTGTA	CTTCTTGTGC	AAGTCAGACT	ACAGCCAACG	ACCCTTACAC	AGTTAGAAGC	1020
	AAAGGCCCTC	ттаатссатс	тсататстса				1050

(45) INFORMATION FOR SEQ ID NO:44:

25

RNSDOCID -WO MODIFICAL! -

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 349 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:
- 30 Met Asn Ser Thr Leu Asp Gly Asn Gln Ser Ser His Pro Phe Cys Leu

ž.,

and the contract of the contra

	1				5			-	-	10					15	
	Leu	Ala	Phe	Gly 20	Tyr	Leu	Glu	Thr	Val 25	Asn	Phe	Cys	Leu	Leu 30	Glu	Val
5	Leu		Ile 35	Val	Phe	Leu	Thr	Val	Leu	Ile	Ile	Ser	Gly 45	Asn	Ile	Ile
	Val	Ile 50	Phe	Val	Phe	His	Cys 55	Ala	Pro	Leu	Leu	Asn 60	His	His	Thr	Thr
	Ser 65	Tyr	Phe	Ile	Gln	Thr 70	Met	Ala	Tyr	Ala	Asp 75	Leu	Phe	Val	Gly	Val 80
10	Ser	Cys	Val	Val	Pro 85	Ser	Leu	Ser	Leu	Leu 90	His	His	Pro	Leu	Pro 95	Val
	Glu	Glu	Ser	Leu 100	Thr	Cys	Gln	Ile	Phe 105	Gly	Phe	Val	Val	Ser 110	Val	Leu
15	Lys	Ser	Val 115	Ser	Met	Ala	Ser	Leu 120	Ala	Cys	Ile	Ser	Ile 125	Asp	Arg	Tyr
	Ile	Ala 130	Ile	Thr	Lys	Pro	Leu 135	Thr	Tyr	Asn	Thr	Leu 140	Val	Thr	Pro	Trp
:	Arg 145	Leu	Arg	Leu	Cys	Ile 150	Phe	Leu	Ile	Trp	Leu 155	Tyr	Ser	Thr	Leu	Val 160
20	Phe	Leu	Pro	Ser	Phe 165	Phe	His	Trp	Gly	Lys 170	Pro	Gly	Tyr	His	Gly 175	Asp
	Val	Phe	Gln	Trp 180	Cys	Ala	Glu	Ser	Trp 185	His	Thr	Asp	Ser	Tyr 190	Phe	Thr
25	Leu	Phe	Ile 195	Val	Met _.	Met	Leu	Tyr 200	Ala	Pro	Ala	Ala	Leu 205	Ile	Val	Cys
	Phe	Thr 210	Tyr	Phe	Asn	Ile	Phe 215	Arg	Ile	Cys	Gln	Gln 220	His	Thr	Lys	Asp
		Ser	Glu	Arg	Gln	Ala 230	Arg	Phe	Ser	Ser	Gln 235	Ser	Gly	Glu	Thr	Gly 240
30	Glu	Val	Gln	Ala	Cys 245	Pro	Asp	Lys	Arg	Tyr 250	Ala	Met	Val	Leu	Phe 255	Arg
	Ile	Thr	Ser	Val 260	Phe	Tyr	Ile	Leu	Trp 265	Leu	Pro	Tyr	Ile	Ile 270	Tyr	Phe
35	Leu	Leu	Glu 275	Ser	Ser	Thr	Gly	His 280	Ser	Asn	Arg	Phe	Ala 285	Ser	Phe	Leu
	Thr	Thr 290	Trp	Leu	Ala	Ile	Ser 295	Asn	Ser	Phe	Cys	Asn 300	Cys	Val	Ile	Tyr

	Ser Leu Ser Asn Ser Val Phe Gln Arg Gly Leu Lys Arg Leu Ser Gl 305 310 315 32	-
	Ala Met Cys Thr Ser Cys Ala Ser Gln Thr Thr Ala Asn Asp Pro Ty 325 330 335	r
5	Thr Val Arg Ser Lys Gly Pro Leu Asn Gly Cys His Ile 340 345	
	(46) INFORMATION FOR SEQ ID NO:45:	
10	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
*	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:	•
15	TCCCCCGGGA AAAAAACCAA CTGCTCCAAA	30
	(47) INFORMATION FOR SEQ ID NO:46:	
20	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic) (xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:	
•	TAGGATCCAT TTGAATGTGG ATTTGGTGAA A	31
25	(48) INFORMATION FOR SEQ ID NO:47:	71
30	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1302 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:	
	ATGTGTTTTT CTCCCATTCT GGAAATCAAC ATGCAGTCTG AATCTAACAT TACAGTGCGA	60
	GATGACATTG ATGACATCAA CACCAATATG TACCAACCAC TATCATATCC GTTAAGCTTT	120
35	CAACTCTCTC TCACCCCATT TCTTATCTTA CAAATTCTCT TCCCACTTCC CACCAACCTC	

The transfer of the second

•	ACTGTATTGG	TACTTTACTG	CATGAAATCC	AACTTAATCA	ACTCTGTCAG	TAACATTATT	240
	ACAATGAATC	TTCATGTACT	TGATGTAATA	ATTTGTGTGG	GATGTATTCC	TCTAACTATA	300
•	GTTATCCTTC	TGCTTTCACT	GGAGAGTAAC	ACTGCTCTCA	TTTGCTGTTT	CCATGAGGCT	360
	TGTGTATCTT	TTGCAAGTGT	CTCAACAGCA	ATCAACGTTT	TTGCTATCAC	TTTGGACAGA	420
5	TATGACATCT	CTGTAAAACC	TGCAAACCGA	ATTCTGACAA	TGGGCAGAGC	TGTAATGTTA	480
	ATGATATCCA	TTTGGATTTT	TTCTTTTTC	TCTTTCCTGA	TTCCTTTTAT	TGAGGTAAAT	540
	TTTTTCAGTC	TTCAAAGTGG	AAATACCTGG	GAAAACAAGA	CACTTTTATG	TGTCAGTACA	600
	AATGAATACT	ACACTGAACT	GGGAATGTAT	TATCACCTGT	TAGTACAGAT	CCCAATATTC	660
	TTTTTCACTG	TTGTAGTAAT	GTTAATCACA	TACACCAAAA	TACTTCAGGC	TCTTAATATT	720
10	CGAATAGGCA	CAAGATTTTC	AACAGGGCAG	AAGAAGAAAG	CAAGAAAGAA	AAAGACAATT	780
	TCTCTAACCA	CACAACATGA	GGCTACAGAC	ATGTCACAAA	GCAGTGGTGG	GAGAAATGTA	840
	GTCTTTGGTG	TAAGAACTTC	AGTTTCTGTA	ATAATTGCCC	TCCGGCGAGC	TGTGAAACGA	900
	CACCGTGAAC	GACGAGAAAG	ACAAAAGAGA	GTCTTCAGGA	TGTCTTTATT	GATTATTTCT	960
	ACATTTCTTC	TCTGCTGGAC	ACCAATTTCT	GTTTTAAATA	CCACCATTTT	ATGTTTAGGC	1020
15	CCAAGTGACC	TTTTAGTAAA	ATTAAGATTG	TGTTTTTTAG	TCATGGCTTA	TGGAACAACT	1080
	ATATTTCACC	СТСТАТТАТА	TGCATTCACT	AGACAAAAAT	TTCAAAAGGT	CTTGAAAAGT	1140
	AAAATGAAAA	AGCGAGTTGT	TTCTATAGTA	GAAGCTGATC	CCCTGCCTAA	TAATGCTGTA	1200
	ATACACAACT	CTTGGATAGA	TCCCAAAAGA	AACAAAAAA	TTACCTTTGA	AGATAGTGAA	1260
	ATAAGAGAAA	AACGTTTAGT	GCCTCAGGTT	GTCACAGACT	AG		1302

- 20 (49) INFORMATION FOR SEQ ID NO:48:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 433 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
- 25 (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

Met Cys Phe Ser Pro Ile Leu Glu Ile Asn Met Gln Ser Glu Ser Asn 1 5 10 15

30 Ile Thr Val Arg Asp Asp Ile Asp Asp Ile Asn Thr Asn Met Tyr Gln 20 25 30

						<i>2.</i>		38	3							
:	Pro		Ser 35	Tyr	Pro	Leu _.		Phe 40	Gln	Val	Ser		Thr 45	Gly	Phe	Leu
	Met	Leu 50	Glu	Ile	Val		Gly 55	Leu	Gly	Ser	Asn	Leu 60	Thr	Val	Leu	Val
5	Leu 65	Tyr	Cys	Met	Lys	Ser 70	Asn	Leu	Ile	Asn	Ser 75	Val	Ser	Asn	Ile	Ile 80
·	Thr	Met	Asn	Leu	His 85	Val	Leu	Asp	Val	Ile 90	Ile	Cys	Val	Gly	Cys 95	Ile
10				100					105		Leu			110		
			115					120			Ser	, , ,	125			
		130					135				Asp	140				
15	145	_				150					Gly 155					160
					165		, .			170	Ser				175	
20				180					185				•	190		Asn
	-		195					200					205			Gly
		210					215					220				Val
25	225					230					235					Ile 240
	_				245					250					255	,
30	-	-		260)				265	5		•		270		Ser
			275	;				280)				285			Val
		290)				295	5				300)		•	Arg
35	305	5				310)				319	5				Ser 320

		•	•		-	. :	325		• .	• .	*, *	330		•	• •	•	335	
			Leu	Cys	Leu	Gly 340	Pro	Ser	Asp	Leu	Leu 345	Val	Lys	Leu	Arg	Leu 350	Cys	Phe
	5		Leu		Met 355	Ala	Tyr	Gly	Thr	Thr 360		Phe	His	Pro	Leu 365	Leu	Tyr	Ala
			Phe	Thr 370	Arg	Gln	Lys	Phe	Gln 375	Lys	Val	Leu	Lys	Ser 380		Met	Lys	Lys
			Arg 385	Val	Val	Ser	Ile	Val 390	Glu	Ala	Asp	Pro	Leu 395		Asn	Asn	Ala	Val 400
	10		Ile	His	Asn	Ser	Trp 405	Ile	Asp	Pro	Lys	Arg 410	Asn	Lys	Lys	Ile	Thr 415	Phe
		i.,	Glu	Asp	Ser	Glu 420	Ile	Arg	Glu	Lys	Arg 425	Leu	Val	Pro		Val 430	Val	Thr
•	15		Asp															
		(50)						•								·		
	20		(i)	(A) (B) (C)	LENC TYI STF	IGTH: PE: r RANDE	30 nucle EDNES	base eic a SS: s	e pai acid singl	rs								
			(ii	.) MC	LECU	TE T	YPE:	: DNZ	A (ge	enomi	Lc)	•						
		• .			EQUEN					SEQ	ID N	10 : 4 9	· :					
	25	GTGA														•		3 (
	30	(51)	(i)	SEQ (A) (B) (C)	UENC LEN TYP STR	E CH IGTH: PE: n	IARAC 31 ucle	DERI Dase eic a	STIC pai cid singl	:S: .rs								
			(ii		LECU					nomi	.c)							
			(xi) SE	QUEN	CE D	ESCR	RIPTI	ON:	SEQ	ID N	io:50):					
		GCAGA	ATTC	c ca	GTGG	CGTG	TTG	TGGT	rgcc	С								31
		(52)	INFO	RMAT	NOI	FOR	SEQ	ID N	10:51	. :								
	35		(i)		UENC LEN						:							

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

The property of the control of the c

- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- 5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

	ATGTTGTGTC	CTTCCAAGAC	AGATGGCTCA	GGGCACTCTG	GTAGGATTCA	CCAGGAAACT	60
	CATGGAGAAG	GGAAAAGGGA	CAAGATTAGC	AACAGTGAAG	GGAGGGAGAA	TGGTGGGAGA	120
	GGATTCCAGA	TGAACGGTGG	GTCGCTGGAG	GCTGAGCATG	CCAGCAGGAT	GTCAGTTCTC	180
	AGAGCAAAGC	CCATGTCAAA	CAGCCAACGC	TTGCTCCTTC	TGTCCCCAGG	ATCACCTCCT	240
10	CGCACGGGGA	GCATCTCCTA	CATCAACATC	ATCATGCCTT	CGGTGTTCGG	CACCATCTGC	300
	CTCCTGGGCA	TCATCGGGAA	CTCCACGGTC	ATCTTCGCGG	TCGTGAAGAA	GTCCAAGCTG	360
	CACTGGTGCA	ACAACGTCCC	CGACATCTTC	ATCATCAACC	TCTCGGTAGT	AGATCTCCTC	420
	TTTCTCCTGG	GCATGCCCTT	CATGATCCAC	CAGCTCATGG	GCAATGGGGT	GTGGCACTTT	480
	GGGGAGACCA	TGTGCACCCT	CATCACGGCC	ATGGATGCCA	ATAGTCAGTT	CACCAGCACC	540
15	TACATCCTGA	CCGCCATGGC	CATTGACCGC	TACCTGGCCA	CTGTCCACCC	CATCTCTTCC	600
	ACGAAGTTCC	GGAAGCCCTC	TGTGGCCACC	CTGGTGATCT	GCCTCCTGTG	GGCCCTCTCC	660
	TTCATCAGCA	TCACCCCTGT	GTGGCTGTAT	GCCAGACTCA	TCCCCTTCCC	AGGAGGTGCA	720
	GTGGGCTGCG	GCATACGCCT	GCCCAACCCA	GACACTGACC	TCTACTGGTT	CACCCTGTAC	780
	CAGTTTTTCC	TGGCCTTTGC	CCTGCCTTTT	GTGGTCATCA	CAGCCGCATA	CGTGAGGATC	840
20	CTGCAGCGCA	TGACGTCCTC	AGTGGCCCCC	GCCTCCCAGC	GCAGCATCCG	GCTGCGGACA	900
	AAGAGGGTGA	CCCGCACAGC	CATCGCCATC	TGTCTGGTCT	TCTTTGTGTG	CTGGGCACCC	960
	TACTATGTGC	TACAGCTGAC	CCAGTTGTCC	ATCAGCCGCC	CGACCCTCAC	CTTTGTCTAC	1020
	TTATACAATG	CGGCCATCAG	CTTGGGCTAT	GCCAACAGCT	GCCTCAACCC	CTTTGTGTAC	1080
	ATCGTGCTCT						1140
25	GGGCAGCTTC	GCGCTGTCAG.	CAACGCTCAG	ACGGCTGACG	AGGAGAGGAC	AGAAAGCAAA	1200
	GGCACCTGA						1209

- (53) INFORMATION FOR SEQ ID NO:52:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 402 amino acids
- 30 (B) TYPE: amino acid

n ye samur ne

	(C) STRANDEDNESS: (D) TOPOLOGY: not relevant																
		(i.	L) MC	LEC	JLE :	TYPE :	pro	oteir	ı		•						
		(x:	i) SI	EQUE	NCE I	DESCI	RIPT	ON:	SEQ	ID 1	10:52	2:					
5		Met 1	Leu	Cys	Pro	Ser 5	Lys	Thr	Asp	Gly	Ser 10	Gly	His	Ser	Gly	Arg 15	Ile
		His	Gln	Glu	Thr 20	His	Gly	Glu	Gly	Lys 25	Arg	Asp	Lys	Ile	Ser 30	Asn	Ser
10		Glu	Gly	Arg 35	Glu	Asn	Gly	Gly	Arg 40	Gly	Phe	Gln	Met	Asn 45	Gly	Gly	Ser
		Leu	Glu 50	Ala	Glu	His	Ala	Ser 55	Arg	Met	Ser	Val	Leu 60	Arg	Ala	Lys	Pro
		Met 65	Ser	Asn	Ser	Gln	Arg 70	Leu	Leu	Leu	Leu	Ser 75	Pro	Gly	Ser	Pro	Pro 80
15		Arg	Thr	Gly	Ser	Ile 85	Ser	Tyr	Ile	Asn	Ile 90	Ile	Met	Pro	Ser	Val 95	Phe ·
	• :	Gly	Thr	Ile	Cys 100	Leu	Leu	Gly	Ile	Ile 105	Gly	Asn	Ser	Thr	Val 110	Ile	Phe
20		Ala	Val	Val 115	Lys	Lys	Ser	Lys	Leu 120	His	Trp	Cys	Asn	Asn 125	Val	Pro	Asp
		Ile	Phe 130	Ile	Ile	Asn	Leu	Ser 135	Val	Val	Asp	Leu	Leu 140	Phe	Leu	Leu	Gly
		Met 145	Pro	Phe	Met	Ile	His 150	Gln	Leu	Met	Gly	Asn 155	Gly	Val	Trp	His	Phe 160
25	•	Gly	Glu	Thr	Met	Cys 165	Thr	Leu	Ile	Thr	Ala 170	Met	Asp	Ala	Asn	Ser 175	Gln
		Phe	Thr	Ser	Thr 180	Tyr	Ile	Leu	Thr	Ala 185	Met	Ala	Ile	Asp	Arg 190	Tyr	Leu
30		Ala	Thr	Val 195	His	Pro	Ile	Ser	Ser 200	Thr	Lys	Phe	Arg	Lys 205	Pro	Ser	Val
		Ala	Thr 210	Leu	Val	Ile	Cys	Leu 215	Leu	Trp	Ala	Leu	Ser 220	Phe	Ile	Ser	Ile
		Thr 225	Pro	Val	Trp	Leu	Tyr 230	Ala	Arg	Leu	Ile	Pro 235	Phe	Pro	Gly	Gly	Ala 240
35		Val	Gly	Cys	Gly	Ile 245	Arg	Leu	Pro	Asn	Pro 250	Asp	Thr	Asp	Leu	Tyr 255	Trp

		Phe	Thr	Leu	Tyr 260	Gln	Phe	Phe	Leu	Ala 265	Phe	Ala	Leu	Pro	Phe 270	Val	Val	
		Ile	Thr	Ala 275	Ala	Tyr	Val	Arg	Ile 280	Leu	Gln	Arg	Met	Thr 235	Ser	Ser	Val	
5		Ala	Pro 290	Ala	Ser	Gln	Arg	Ser 295	Ile	Arg	Leu	Arg	Thr 300	Lys	Arg	Val	Thr	
		Arg 305	Thr	Ala	Ile	Ala	Ile 310	Cys	Leu	Val	Phe	Phe 315	Val	Cys	Trp	Ala	Pro 320	
10		Tyr	Tyr	Val	Leu	Gln 325	Leu	Thr	Gln	Leu	Ser 330	Ile	Ser	Arg	Pro	Thr 335	Leu	
	* :	Thr	Phe	Val	Tyr 340	Leu	Tyr	Asn	Ala	Ala 345	Ile	Ser	Leu	Gly	Tyr 350	Ala	Asn	
		Ser	Cys	Leu 355	Asn	Pro	Phe	Val	Tyr 360	Ile	Val	Leu	Cys	Glu 365	Thr	Phe	Arg	
15		Lys	Arg 370	Leu	Val	Leu	Ser	Val 375	Lys	Pro	Ala	Ala	Gln 380	Gly	Gln	Leu	Arg	
		Ala 385	Val	Ser	Asn	Ala	Gln 390	Thr	Ala	Asp	Glu	Glu 395	Arg	Thr	Glu	Ser	Lys 400	
20		Gly	Thr							•			•	٠	•	•		
	(54)	INF	ORMA!	TION	FOR	SEQ	ID 1	NO : 53	3:									
25		(i)	(B)	LEI TYI	CE CH NGTH: PE: 1 RANDI POLO(: 27 nucle EDNES	base eic a SS: s	e pa: acid sing:	irs						`,			
		(i:	i) M	OLECT	JLE :	TYPE:	: DN	A (ge	enom:	ic)						•		
		(x:	i) SI	EQUEI	NCE I	DESCI	RIPT	ION:	SEQ	ID N	10:53	3:						
	GGCG	GATC	CA TO	GGAT	GTGA	TTC	CCA	A.										27
30	(55)	INF	ORMA!	LION	FOR	SEQ	ID 1	NO : 54	1 :				•			,		
35		(i)	(B)	LEI TYI	CE CI NGTH: PE: 1 RANDI POLO	: 27 nucle EDNES	base eic a SS: a	e pa: acid sing:	irs								·	
		(i:	i) M	OLECT	ULE :	TYPE:	: DN	A (ge	enom:	ic)								
		(x:	i) si	EOUE	NCE I	DESCI	RIPT	ION:	SEO	ID 1	NO : 54	4 :						

PCT/US99/23938

5

43

\$ 400

GGCGGATCCC	TACA	CGGCAC	TGCTGAA

27

arta in a

(56) INFORMATION FOR SEQ ID NO:55:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1128 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

ATGGATGTGA CTTCCCAAGC CCGGGGCGTG GGCCTGGAGA TGTACCCAGG CACCGCGCAC 60 10 GCTGCGGCCC CCAACACCAC CTCCCCCGAG CTCAACCTGT CCCACCCGCT CCTGGGCACC GCCCTGGCCA ATGGGACAGG TGAGCTCTCG GAGCACCAGC AGTACGTGAT CGGCCTGTTC 180 CTCTCGTGCC TCTACACCAT CTTCCTCTTC CCCATCGGCT TTGTGGGCAA CATCCTGATC 240 CTGGTGGTGA ACATCAGCTT CCGCGAGAAG ATGACCATCC CCGACCTGTA CTTCATCAAC 300 CTGGCGGTGG CGGACCTCAT CCTGGTGGCC GACTCCCTCA TTGAGGTGTT CAACCTGCAC 360 GAGCGGTACT ACGACATCGC CGTCCTGTGC ACCTTCATGT CGCTCTTCCT GCAGGTCAAC ATGTACAGCA GCGTCTTCTT CCTCACCTGG ATGAGCTTCG ACCGCTACAT CGCCCTGGCC AGGGCCATGC GCTGCAGCCT GTTCCGCACC AAGCACCACG CCCGGCTGAG CTGTGGCCTC ATCTGGATGG CATCCGTGTC AGCCACGCTG GTGCCCTTCA CCGCCGTGCA CCTGCAGCAC 600 ACCGACGAGG CCTGCTTCTG TTTCGCGGAT GTCCGGGAGG TGCAGTGGCT CGAGGTCACG 20 CTGGGCTTCA TCGTGCCCTT CGCCATCATC GGCCTGTGCT ACTCCCTCAT TGTCCGGGTG 720 CTGGTCAGGG CGCACCGGCA CCGTGGGCTG CGGCCCCGGC GGCAGAAGGC GCTCCGCATG GTGCACCTCC TGCAGCGGAC GCAGCCTGGG GCCGCTCCCT GCAAGCAGTC TTTCCGCCAT 900 GCCCACCCC TCACGGGCCA CATTGTCAAC CTCGCCGCCT TCTCCAACAG CTGCCTAAAC 25 CCCCTCATCT ACAGCTTTCT CGGGGAGACC TTCAGGGACA AGCTGAGGCT GTACATTGAG 1020 CAGAAAACAA ATTTGCCGGC CCTGAACCGC TTCTGTCACG CTGCCCTGAA GGCCGTCATT 1080 CCAGACAGCA CCGAGCAGTC GGATGTGAGG TTCAGCAGTG CCGTGTGA 1128

(57) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 375 amino acids

(B) TYPE: amino acid(C) STRANDEDNESS:(D) TOPOLOGY: not relevant																
5	(ii	L) MC	DLECT	TLE 3	TYPE:	pro	oteir	1								
	(xi	i) SI	EQUE1	ICE I	DESCI	RIPTI	ON:	SEQ	ID 1	NO : 56	5:					
	Met 1	Asp	Val	Thr	Ser 5	Gln	Ala	Arg	Gly	Val	Gly	Leu	Glu	Met	Tyr 15	Pro
10	Gly	Thr	Ala	His 20	Ala	Ala	Ala	Pro	Asn 25	Thr	Thr	Ser	Pro	Glu 30	Leu	Asn
	Leu	Ser	His 35	Pro	Leu	Leu	Gly	Thr 40	Ala	Leu	Ala	Asn	Gly 45	Thr	Gly	Glu
•	Leu	Ser 50	Glu	His	Gln	Gln	Tyr 55	Val	Ile	Gly	Leu	Phe 60	Leu	Ser	Cys	Leu
15	Tyr 65	Thr	Ile	Phe	Leu	Phe 70	Pro	Ile	Gly	Phe	Val 75	Gly	Asn	Ile	Leu	Ile 80
	Leu	Val	Val	Asn	Ile 85	Ser	Phe	Arg	Glu	Lys 90	Met	Thr	Ile	Pro	Asp 95	Leu
20	Tyr	Phe	Ile	Asn 100	Leu	Ala	Val	Ala	Asp 105	Leu	Ile	Leu	Val	Ala 110	Asp	Ser
	Leu	Ile	Glu 115	Val	Phe	Asn	Leu	His 120	Glu	Arg	Tyr	Tyr	Asp 125	Ile	Ala	Val
	Leu	Cys 130	Thr	Phe	Met	Ser	Leu 135	Phe	Leu	Gln	Val	Asn 140	Met	Tyr	Ser	Ser
25	Val 145	Phe	Phe	Leu	Thr	Trp 150	Met	Ser	Phe	Asp	Arg 155	Tyr	Ile	Ala	Leu	Ala 160
	Arg	Ala	Met	Arg	Cys 165	Ser	Leu	Phe	Arg	Thr 170	Lys	His	His	Ala	Arg 175	Leu
30	Ser	Cys	Gly	Leu 180	Ile	Trp	Met	Ala	Ser 185	Val	Ser	Ala	Thr	Leu 190	Val	Pro
	Phe	Thr	Ala 195	Val	His	Leu	Gln	His 200	Thr	Asp	Glu	Ala	Cys 205	Phe	Cys	Phe
	Ala	Asp 210	Val	Arg	Glu	Val	Gln 215	Trp	Leu	Glu	Val	Thr 220	Leu	Gly	Phe	Ile
35	Val 225	Pro	Phe	Ala	Ile	Ile 230	Gly	Leu	Cys	Tyr	Ser 235	Leu	Ile	Val	Arg	Val 240
	Leu	Val	Arg	Ala	His	Arg	His	Arg	Gly	Leu	Arg	Pro	Arg	Arg	Gln	Lys

PCT/US99/23938

						245				*	250	Į.				255	
		Ala	Leu	Arg	Met 260	Ile	Leu	Ala	Val	Val 265	Leu	Val	Phe	Phe	Val 270	Cys	Trp
5		Leu	Pro	Glu 275	Asn	Val	Phe	Ile	Ser 280	Val	His	Leu	Leu	Gln 285	Arg	Thr	Gln
		Pro	Gly 290	Ala	Ala	Pro	Cys	Lys 295	Gln	Ser	Phe	Arg	His 300	Ala	His	Pro	Leu
		Thr 305	Gly	His	Ile	Val	Asn 310	Leu	Ala	Ala	Phe	Ser 315	Asn	Ser	Cys	Leu	Asn 320
10		Pro	Leu	Ile	Tyr	Ser 325	Phe	Leu	Gly	Glu	Thr 330	Phe	Arg	Asp	Lys	Leu 335	Arg
		Leu				Gln					Pro	Ala	Leu	Asn	Arg 350	Phe	Cys
15		His	Ala	Ala 355	Leu	Lys	Ala	Val	Ile 360	Pro	Asp	Ser	Thr	Glu 365	Gln	Ser	Asp
		Val	Arg 370	Phe	Ser	Ser	Ala	Val 375								ē	
	(58)	INFO	RMAI	rion	FOR	SEQ	ID 1	NO : 57	7:				~				
20			(A) (B) (C) (D)	LEN TYI STI TOI	GTH: PE: I RANDI POLOG	HARAC : 31 nucle EDNES EY:]	base eic a SS: s linea	e pai acid sing] ar	irs le	ic)					÷		
25						DESCE					NO : 57	7:					
	AAGGA	ATTC	A CO	GCCG	GGTC	ATC	CCA	rtcc	C			•					3
	(59)	INFO	RMAT	CION	FOR	SEQ	ID 1	10:58	B:								
30		(i)	(A) (B) (C)	LEN TYP STF	IGTH: PE: r RANDE	HARAC : 30 nucle EDNES EY: 1	base ic a SS: s	e pai acid singl	rs								
		(ii	.) MC	LECU	LE 1	TYPE:	DN	A (ge	enomi	ic)							
		(xi) SE	QUEN	ICE I	DESC	RIPT	ON:	SEQ	ID 1	10 : 58	3:					
35	GGTGG	ATCO	A TA	AACA	CGGG	G CG1	TGAC	GAC									3
	(60)	INFO	RMAT	CION	FOR	SEQ	ID 1	NO : 59) :					-	:		

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 960 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
- 5 (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

ATGCCATTCC CAAACTGCTC AGCCCCCAGC ACTGTGGTGG CCACAGCTGT GGGTGTCTTG 60 CTGGGGCTGG AGTGTGGGCT GGGTCTGCTG GGCAACGCGG TGGCGCTGTG GACCTTCCTG 120 10 TTCCGGGTCA GGGTGTGGAA GCCGTACGCT GTCTACCTGC TCAACCTGGC CCTGGCTGAC 180 CTGCTGTTGG CTGCGTGCCT GCCTTTCCTG GCCGCCTTCT ACCTGAGCCT CCAGGCTTGG 240 CATCTGGGCC GTGTGGGCTG CTGGGCCCTG CGCTTCCTGC TGGACCTCAG CCGCAGCGTG 300 GGGATGGCCT TCCTGGCCGC CGTGGCTTTG GACCGGTACC TCCGTGTGGT CCACCCTCGG 360 CTTAAGGTCA ACCTGCTGTC TCCTCAGGCG GCCCTGGGGG TCTCGGGCCT CGTCTGGCTC 420 15 CTGATGGTCG CCCTCACCTG CCCGGGCTTG CTCATCTCTG AGGCCGCCCA GAACTCCACC 480 AGGTGCCACA GTTTCTACTC CAGGGCAGAC GGCTCCTTCA GCATCATCTG GCAGGAAGCA 540 CTCTCCTGCC TTCAGTTTGT CCTCCCCTTT GGCCTCATCG TGTTCTGCAA TGCAGGCATC 600 ATCAGGGCTC TCCAGAAAAG ACTCCGGGAG CCTGAGAAAC AGCCCAAGCT TCAGCGGGCC 660 CAGGCACTGG TCACCTTGGT GGTGGTGCTG TTTGCTCTGT GCTTTCTGCC CTGCTTCCTG 720 20 GCCAGAGTCC TGATGCACAT CTTCCAGAAT CTGGGGAGCT GCAGGGCCCT TTGTGCAGTG 780 GCTCATACCT CGGATGTCAC GGGCAGCCTC ACCTACCTGC ACAGTGTCGT CAACCCCGTG 840 GTATACTGCT TCTCCAGCCC CACCTTCAGG AGCTCCTATC GGAGGGTCTT CCACACCCTC 900 CGAGGCAAAG GGCAGGCAGC AGAGCCCCCA GATTTCAACC CCAGAGACTC CTATTCCTGA 960

- (61) INFORMATION FOR SEQ ID NO:60:
- 25 (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 319 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
- 30 (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

Met Pro Phe Pro Asn Cys Ser Ala Pro Ser Thr Val Val Ala Thr Ala

בובחרות שות החוזימה ו

	1				5		· · ·		*	10					15	
	Val	Gly	Val	Leu 20	Leu	Gly	Leu	Glu	Cys 25	Gly	Leu	Gly	Leu	Leu 30	Gly	Asn
5	Ala	Val	Ala 35	Leu	Trp	Thr	Phe	Leu 40	Phe	Arg	Val	Arg	Val 45	Trp	Lys	Pro
	туr	Ala 50	Val	Tyr	Leu	Leu	Asn 55	Leu	Ala	Leu	Ala	Asp 60	Leu	Leu	Leu	Ala
	Ala 65	Cys	Leu	Pro	Phe	Leu 70	Ala	Ala	Phe	Tyr	Leu 75	Ser	Leu	Gln	Ala	Trp 80
10	His	Leu	Gly	Arg	Val 85	Gly	Cys	Trp	Ala	Leu 90	Arg	Phe	Leu	Leu	Asp 95	Leu
. 4	Ser	Arg	Ser	Val 100	Gly	Met	Ala	Phe	Leu 105	Ala	Ala	Val	Ala	Leu 110	Asp	Arg
15	Tyr	Leu	Arg 115	Val	Val	His	Pro	Arg 120	Leu	Lys	Val	Asn	Leu 125	Leu	Ser	Pro
	Gln	Ala 130	Ala	Leu	Gly	Val	Ser 135	Gly	Leu	Val	Trp	Leu 140	Leu	Met	Val	Ala
	Leu 145	Thr	Cys	Pro	Gly	Leu 150	Leu	Ile	Ser	Glu	Ala 155	Ala	Gln	Asn	Ser	Thr 160
20	Arg	Cys	His	Ser	Phe 165	Tyr	Ser	Arg	Ala	Asp 170	Gly	Ser	Phe	Ser	Ile 175	Ile
	Trp	Gln	Glu	Ala 180	Leu	Ser	Cys	Leu	Gln 185	Phe	Val	Leu	Pro	Phe 190	Gly	Leu
25	Ile	Val	Phe 195	Cys	Asn	Ala	Gly	Ile 200	Ile	Arg	Ala	Leu	Gln 205	Lys	Arg	Leu
	Arg	Glu 210	Pro	Glu	Lys	Gln	Pro 215	Lys	Leu	Gln	Arg	Ala 220	Gln	Ala	Leu	Val
	Thr 225	Leu	Val	Val	Val	Leu 230	Phe	Ala	Leu	Cys	Phe 235	Leu	Pro	Cys	Phe	Leu 240
30	Ala	Arg	Val	Leu	Met 245	His	Ile	Phe	Gln	Asn 250	Leu	Gly	Ser	Cys	Arg 255	Ala
	Leu	Сув	Ala	Val 260	Ala	His	Thr	Ser	Asp 265	Val	Thr	Gly	Ser	Leu 270	Thr	Tyr
35	Leu	His	Ser 275	Val	Val	Asn	Pro	Val 280	Val	Tyr	Cys	Phe	Ser 285	Ser	Pro	Thr
	Phe	Arg 290	Ser	Ser	Tyr	Arg	Arg 295	Val	Phe	His	Thr	Leu 300	Arg	Gly	Lys	Gly

Gln Ala Ala Glu Pro Pro Asp Phe Asn Pro Arg Asp Ser Tyr Ser 305 310 315

(62) INFORMATION FOR SEQ ID NO:61:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1143 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- 10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

	ATGGAGGAAG	GTGGTGATTT	TGACAACTAC	TATGGGGCAG	ACAACCAGTC	TGAGTGTGAG	60
	TACACAGACT	GGAAATCCTC	GGGGCCCTC	ATCCCTGCCA	TCTACATGTT	GGTCTTCCTC	120
	CTGGGCACCA	CGGGAAACGG	TCTGGTGCTC	TGGACCGTGT	TTCGGAGCAG	CCGGGAGAAG	180
	AGGCGCTCAG	CTGATATCTT	CATTGCTAGC	CTGGCGGTGG	CTGACCTGAC	CTTCGTGGTG	240
15	ACGCTGCCCC	TGTGGGCTAC	CTACACGTAC	CGGGACTATG	ACTGGCCCTT	TGGGACCTTC	300
	TTCTGCAAGC	TCAGCAGCTA	CCTCATCTTC	GTCAACATGT	ACGCCAGCGT	CTTCTGCCTC	360
	ACCGGCCTCA	GCTTCGACCG	CTACCTGGCC	ATCGTGAGGC	CAGTGGCCAA	TGCTCGGCTG	420
	AGGCTGCGGG	TCAGCGGGGC	CGTGGCCACG	GCAGTTCTTT	GGGTGCTGGC	CGCCCTCCTG	480
	GCCATGCCTG	TCATGGTGTT	ACGCACCACC	GGGGACTTGG	AGAACACCAC	TAAGGTGCAG	540
20	TGCTACATGG	ACTACTCCAT	GGTGGCCACT	GTGAGCTCAG	AGTGGGCCTG	GGAGGTGGGC	600
	CTTGGGGTCT	CGTCCACCAC	CGTGGGCTTT	GTGGTGCCCT	TCACCATCAT	GCTGACCTGT	660
	TACTTCTTCA	TCGCCCAAAC	CATCGCTGGC	CACTTCCGCA	AGGAACGCAT	CGAGGGCCTG	720
	CGGAAGCGGC	GCCGGCTGCT	CAGCATCATC	GTGGTGCTGG	TGGTGACCTT	TGCCCTGTGC	780
	TGGATGCCCT	ACCACCTGGT	GAAGACGCŢG	TACATGCTGG	GCAGCCTGCT	GCACTGGCCC	840
25	TGTGACTTTG	ACCTCTTCCT	CATGAACATC	TTCCCCTACT	GCACCTGCAT	CAGCTACGTC	900
	AACAGCTGCC	TCAACCCCTT	CCTCTATGCC	TTTTTCGACC	CCCGCTTCCG	CCAGGCCTGC	960
	ACCTCCATGC	TCTGCTGTGG	CCAGAGCAGG	TGCGCAGGCA	CCTCCCACAG	CAGCAGTGGG	1020
	GAGAAGTCAG	CCAGCTACTC	TTCGGGGCAC	AGCCAGGGGC	CCGGCCCCAA	CATGGGCAAG	1080
	GGTGGAGAAC	AGATGCACGA	GAAATCCATC	CCCTACAGCC	AGGAGACCCT	TGTGGTTGAC	1140
30	TAG						1143

(63) INFORMATION FOR SEQ ID NO:62:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 380 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:

5

- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein

	(x:	i) Si	EQUE	NCE 1	DESCI	RIPT	ION:	SEQ	ID 1	NO:6	2:		•			
10	Met 1	Glu	Glu	Gly	Gly 5	Asp	Phe	Asp	Asn	Туr 10	Tyr	Gly	Ala	Asp	Asn 15	Gln
	Ser	Glu	Cys	Glu 20		Thr	Asp	Trp	Lys 25	Ser	Ser	Gly	Ala	Leu 30	Ile	Pro
	Ala	Ile	Tyr 35	Met	Leu	Val	Phe	Leu 40	Leu	Gly	Thr	Thr	Gly 45	Asn	Gly	Leu
15	Val	Leu 50	Trp	Thr	Val	Phe	Arg 55	Ser	Ser	Arg	Glu	Lys 60	Arg	Arg	Ser	Ala
	Asp 65	Ile	Phe	Ile	Ala	Ser 70	Leu	Ala	Val	Ala	Asp 75	Leu	Thr	Phe	Val	Val 80
20	Thr	Leu	Pro	Leu	Trp 85	Ala	Thr	Tyr	Thr	Tyr 90	Arg	Asp	Tyr	Asp	Trp 95	Pro
	Phe	Gly	Thr	Phe 100	Phe	Cys	Lys	Leu	Ser 105	Ser	Tyr	Leu	Ile	Phe 110	Val	Asn
	Met	Tyr	Ala 115	Ser	Val	Phe	Cys	Leu 120	Thr	Gly	Leu	Ser	Phe 125	Asp	Arg	Tyr
25	Leu	Ala 130	Ile	Val	Arg	Pro	Val 135	Ala	Asn	Ala	Arg	Leu 140	Arg	Leu	Arg	Val
	Ser 145	Gly	Ala	Val	Ala	Thr 150	Ala	Val	Leu	Trp	Val 155	Leu	Ala	Ala	Leu	Leu 160
30	Ala	Met	Pro	Val	Met 165	Val	Leu	Arg	Thr	Thr 170	Gly	Asp	Leu	Glu	Asn 175	Thr
	Thr	Lys	Val	Gln 180	Cys	Tyr	Met	Asp	Tyr 185	Ser	Met	Val	Ala	Thr 190	Val	Ser
	Ser	Glu	Trp 195	Ala	Trp	Glu	Val	Gly 200	Leu	Gly	Val	Ser	Ser 205	Thr	Thr	Val
35	Gly	Phe 210	Val	Val	Pro	Phe	Thr 215	Ile	Met	Leu	Thr	Cys 220	Tyr	Phe	Phe	Ile

Ala Gln Thr Ile Ala Gly His Phe Arg Lys Glu Arg Ile Glu Gly Leu

		225					230		٠٠.			235		•	•	•	240	
		Arg	Lys	Arg	Arg	Arg 245	Leu	Leu	Ser	Ile	Ile 250	Val	Val	Leu	Val	Val 255	Thr	•
5		Phe	Ala	Leu	Cys 260	Trp	Met	Pro	Tyr	His 265	Leu	Val	Lys	Thr	Leu 270	Tyr	Met	
		Leu	Gly	Ser 275	Leu	Leu	His	Trp	Pro 280	Cys	Asp	Phe	Asp	Leu 285	Phe	Leu	Met	
		Asn	Ile 290	Phe	Pro	Tyr	Cys	Thr 295	Cys	Ile	Ser	Tyr	Val 300	Asn	Ser	Cys	Leu	
10		Asn 305	Pro	Phe	Leu	Tyr	Ala 310	Phe	Phe	Asp	Pro	Arg 315	Phe	Arg	Gln	Ala	Cys 320	
. •		Thr	Ser	Met	Leu	Cys 325	Cys	Gly	Gln	Ser	Arg 330	Cys	Ala	Gly	Thr	Ser 335	His	
15		Ser	Ser	Ser	Gly 340	Glu	Lys	Ser	Ala	Ser 345	Tyr	Ser.	Ser	Gly	His 350	Ser	Gln	
		Gly	Pro	Gly 355	Pro	Asn	Met	Gly	Lys 360	Gly	Gly	Glu	Gln	Met 365	His	Glu	Lys	
•		Ser	Ile 370	Pro	Tyr	Ser	Gln	Glu 375	Thr	Leu	Val	Val	Asp 380		*.			
20	(64)	INFO	RMAT	CION	FOR	SEQ	ID N	iO:63	:			;						
		(i)		UENC LEN														
				TYP STR					.e									
25				TOP														
		(ii) MC	LECU	LE I	YPE:	DNA	(ge	nomi	.c)		•	• •					
		(xi) SE	QUEN	CE D	ESCR	IPTI	ON:	SEQ	ID N	10:63	:						
	TGAGA	ATTC	T GG	TGAC	TCAC	AGC	CGGC	ACA	G									31
	(65)	INFO	RMAT	'ION	FOR	SEQ	ID N	0:64	:							•		
30		(i)		UENC														•
			(B) (C)	TYP STR	E: n	ucle DNES	ic a S: s	cid ingl										
35		,,,		TOP														
33				LECU														
		(Xi) SE	QUEN	CE D	ESCR	IPTI.	ON:	SEQ	ID N	0:64	:		•				

51

GCCGGATCCA AGGAAAAGCA GCAATAAAAG G

31

(66) INFORMATION FOR SEQ ID NO:65:

5

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1119 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

10	ATGAACTACC	CGCTAACGCT	GGAAATGGAC	CTCGAGAACC	TGGAGGACCT	GTTCTGGGAA	60
	CTGGACAGAT	TGGACAACTA	TAACGACACC	TCCCTGGTGG	AAAATCATCT	CTGCCCTGCC	120
	ACAGAGGGTC	CCCTCATGGC	CTCCTTCAAG	GCCGTGTTCG	TGCCCGTGGC	CTACAGCCTC	180
	ATCTTCCTCC	TGGGCGTGAT	CGGCAACGTC	CTGGTGCTGG	TGATCCTGGA	GCGGCACCGG	240
	CAGACACGCA	GTTCCACGGA	GACCTTCCTG	TTCCACCTGG	CCGTGGCCGA	CCTCCTGCTG	300 0
15	GTCTTCATCT	TGCCCTTTGC	CGTGGCCGAG	GGCTCTGTGG	GCTGGGTCCT	GGGGACCTTC	360
	CTCTGCAAAA	CTGTGATTGC	CCTGCACAAA	GTCAACTTCT	ACTGCAGCAG	CCTGCTCCTG	420
	GCCTGCATCG	CCGTGGACCG	CTACCTGGCC	ATTGTCCACG	CCGTCCATGC	CTACCGCCAC	480
	CGCCGCCTCC	TCTCCATCCA	CATCACCTGT	GGGACCATCT	GGCTGGTGGG	CTTCCTCCTT	540
	GCCTTGCCAG	AGATTCTCTT	CGCCAAAGTC	AGCCAAGGCC	ATCACAACAA	CTCCCTGCCA	600
20	CGTTGCACCT	TCTCCCAAGA	GAACCAAGCA	GAAACGCATG	CCTGGTTCAC	CTCCCGATTC	660
	CTCTACCATG	TGGCGGGATT	CCTGCTGCCC	ATGCTGGTGA	TGGGCTGGTG	CTACGTGGGG	720
	GTAGTGCACA	GGTTGCGCCA	GGCCCAGCGG	CGCCCTCAGC	GGCAGAAGGC	AGTCAGGGTG	780
	GCCATCCTGG	TGACAAGCAT	CTTCTTCCTC	TGCTGGTCAC	CCTACCACAT	CGTCATCTTC	840
	CTGGACACCC	TGGCGAGGCT	GAAGGCCGTG	GACAATACCT	GCAAGCTGAA	TGGCTCTCTC	900
25	CCCGTGGCCA	TCACCATGTG	TGAGTTCCTG	GGCCTGGCCC	ACTGCTGCCT	CAACCCCATG	960
	CTCTACACTT	TCGCCGGCGT	GAAGTTCCGC	AGTGACCTGT	CGCGGCTCCT	GACCAAGCTG	1020
	GGCTGTACCG	GCCCTGCCTC	CCTGTGCCAG	CTCTTCCCTA	GCTGGCGCAG	GAGCAGTCTC	1080
	TCTGAGTCAG	AGAATGCCAC	CTCTCTCACC	ACGTTCTAG			1119

(67) INFORMATION FOR SEQ ID NO:66:

30 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 372 amino acids

(B) TYPE: amino acid

			TOP				relev	ant								
5	(ii	L) MC	DLECT	JLE 1	YPE:	pro	oteir	1								
	(xi	l) SI	EQUEN	ICE I	ESCI	RIPT	ON:	SEQ	ID 1	NO : 6	5 :	. ,			,	
	Met 1	Asn	Tyr	Pro	Leu 5	Thr	Leu	Glu	Met	Asp 10	Leu	Glu	Asn	Leu	Glu 15	Asp
10	Leu	Phe	Trp	Glu 20	Leu	Asp	Arg	Leu	Asp 25	Asn	Tyr	Asn	Asp	Thr 30	Ser	Leu
	Val	Glu	Asn 35	His	Leu	Cys	Pro	Ala 40	Thr	Glu	Gly	Pro	Leu 45	Met	Ala	Ser
	Phe	Lys 50	Ala	Val	Phe	Val	Pro 55	Val	Ala	Tyr	Ser	Leu 60	Ile	Phe	Leu	Leu
15	Gly 65	Val	Ile	Gly	Asn	Val 70	Leu	Val	Leu	Val	Ile 75	Leu	Glu	Arg	His	Arg 80
	Gln	Thr	Arg	Ser	Ser 85	Thr	Glu	Thr	Phe	Leu 90	Phe	His	Leu	Ala	Val 95	Ala
20	Asp	Leu	Leu	Leu 100	Val	Phe	Ile	Leu	Pro 105	Phe	Ala	Val	Ala	Glu 110	Gly	Ser
	Val	Gly	Trp 115	Val	Leu	Gly	Thr	Phe 120	Leu	Cys	Lys	Thr	Val 125	Ile	Ala	Leu
	His	Lys 130	Val	Asn	Phe	Tyr	Cys 135		Ser	Leu	Leu	Leu 140	Ala	Cys	Ile	Ala
25	Val 145	Asp	Arg	Tyr	Leu	Ala 150	Ile	Val	His	Ala	Val 155	His	Ala	Tyr	Arg	His 160
	Arg	Arg	Leu	Leu	Ser 165	Ile	His	Ile	Thr	Cys 170	Gly	Thr	Ile	Trp	Leu 175	Val
30	Gly	Phe	Leu	Leu 180	Ala	Leu	Pro-	Glu	Ile 185	Leu	Phe	Ala	Lys	Val 190	Ser	Gln
	Gly	His	His 195	Asn	Asn	Ser	Leu	Pro 200	Arg	Cys	Thr	Phe	Ser 205	Gln	Glu	Asn
	Gln	Ala 210	Glu	Thr	His	Ala	Trp 215	Phe	Thr	Ser	Arg	Phe 220	Leu	Tyr	His	Val
35	Ala 225	Gly	Phe	Leu	Leu	Pro 230	Met	Leu	Val	Met	Gly 235	Trp	Cys	Tyr	Val	Gly 240
	17a]	Val	Hie	Ara	T.en:	7.~~	Glr	λla	Glr	7.2~	7~~	Dro	GJ ~	7 ÷-	C1=	T

בו ואמנינינים האות האותרושות

						245		•.:			250					255		
		Ala	Val	Arg	Val 260	Ala	Ile	Leu	Val	Thr 265	Ser	Ile	Phe	Phe	Leu 270	Cys	Trp	
5		Ser	Pro	Tyr 275	His	Ile	Val	Ile	Phe 280	Leu	Asp	Thr	Leu	Ala 285	Arg	Leu	Lys	
		Ala	Val 290	Asp	Asn	Thr	Cys	Lys 295	Leu	Asn	Gly	Ser	Leu 300	Pro		Ala	Ile	
		Thr 305	Met	Cys	Glu	Phe	Leu 310	Gly	Leu	Ala	His	Cys 315	Cys	Leu	Asn	Pro	Met 320	
10		Leu	туг	Thr	Phe	Ala 325	Gly	Val	Lys	Phe	Arg 330	Ser	Asp	Leu	Ser	Arg 335	Leu	
	• *		Thr	Lys	Leu 340	Gly	Cys	Thr	Gly	Pro 345	Ala	Ser	Leu	Cys	Gln 350	Leu	Phe	
15		Pro	Ser	Trp 355	Arg	Arg	Ser	Ser	Leu 360	Ser	Glu	Ser	Glu	Asn 365	Ala	Thr	Ser	
	<i>;</i> . ·	Leu	Thr 370	Thr	Phe										•			
	(68)	INFO	ORMA:	rion	FOR	SEQ	ID 1	10:67	7:									
20		(i)	(B)	LEN TYP	NGTH: PE: 1 RANDI	: 30 nucle EDNES	base eic a SS: s	e pai acid singl	irs									
		(ii	i) Mo	DLEC	JLE T	YPE:	DNA	A (ge	nomi	ic)								
25		(x:	i) SI	EQUE1	ICE I	ESCF	RIPTI	ON:	SEQ	ID 1	10:67	7:						٠.
	CAAA	CTT	A AE	AGCTO	CAC	GTO	CAG	AGAC										30
	(69)	INFO	ORMA!	rion	FOR	SEQ	ID 1	10:68	3:									
30		(i·)	(B)	LEN TYP	IGTH: PE: 1 RANDI	: 30 nucle EDNES	base ic a SS: s	e pai acid singl	irs			:						
		(i:	i) Mo	OLECT	TLE T	TYPE :	DN2	A (ge	enomi	ic)								
		(x:	i) SI	EQUE	ICE I	DESC	RIPT	ON:	SEQ	ID 1	10:68	3:						
35	GCGG	ATCC	CG AC	STCAC	CACCO	TGC	CTG	GCC										30
	(70)	INFO	ORMA:	rion	FOR	SEQ	ID 1	10:69) :		,						•	

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1128 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5

- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

	ATGGATGTGA	CTTCCCAAGC	CCGGGGCGTG	GGCCTGGAGA	TGTACCCAGG	CACCGCGCAG	60
	CCTGCGGCCC	CCAACACCAC	CTCCCCGAG	CTCAACCTGT	CCCACCCGCT	CCTGGGCACC	120
10	GCCCTGGCCA	ATGGGACAGG	TGAGCTCTCG	GAGCACCAGC	AGTACGTGAT	CGGCCTGTTC	180
	CTCTCGTGCC	TCTACACCAT	CTTCCTCTTC	CCCATCGGCT	TTGTGGGCAA	CATCCTGATC	240
	CTGGTGGTGA	ACATCAGCTT	CCGCGAGAAG	ATGACCATCC	CCGACCTGTA	CTTCATCAAC	300
	CTGGCGGTGG	CGGACCTCAT	CCTGGTGGCC	GACTCCCTCA	TTGAGGTGTT	CAACCTGCAC	360
	GAGCGGTACT	ACGACATCGC	CGTCCTGTGC	ACCTTCATGT	CGCTCTTCCT	GCAGGTCAAC	420
15	ATGTACAGCA	GCGTCTTCTT	CCTCACCTGG	ATGAGCTTCG	ACCGCTACAT	CGCCCTGGCC	480
	AGGGCCATGC	GCTGCAGCCT	GTTCCGCACC	AAGCACCACG	CCCGGCTGAG	CTGTGGCCTC	540
	ATCTGGATGG	CATCCGTGTC	AGCCACGCTG	GTGCCCTTCA	CCGCCGTGCA	CCTGCAGCAC	600
	ACCGACGAGG	CCTGCTTCTG	TTTCGCGGAT	GTCCGGGAGG	TGCAGTGGCT	CGAGGTCACG	660
	CTGGGCTTCA	TCGTGCCCTT	CGCCATCATC	GGCCTGTGCT	ACTCCCTCAT	TGTCCGGGTG	720
20	CTGGTCAGGG	CGCACCGGCA	CCGTGGGCTG	CGGCCCCGGC	GGCAGAAGGC	GCTCCGCATG	780
	ATCCTCGCGG	TGGTGCTGGT	CTTCTTCGTC	TGCTGGCTGC	CGGAGAACGT	CTTCATCAGC	840
	GTGCACCTCC	TGCAGCGGAC	GCAGCCTGGG	GCCGCTCCCT	GCAAGCAGTC	TTTCCGCCAT	900
	GCCCACCCC	TCACGGGCCA	CATTGTCAAC	CTCACCGCCT	TCTCCAACAG	CTGCCTAAAC	960
	CCCCTCATCT	ACAGCTTTCT	CGGGGAGACC	TTCAGGGACA	AGCTGAGGCT	GTACATTGAG	1020
25	CAGAAAACAA	ATTTGCCGGC	CCTGAACCGC	TTCTGTCACG	CTGCCCTGAA	GGCCGTCATT	1080
	CCAGACAGCA	CCGAGCAGTC	GGATGTGAGG	TTCAGCAGTG	CCGTGTAG		1128

(71) INFORMATION FOR SEQ ID NO:70:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 375 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant

		(i	i) M	OLEC	ULE '	TYPE	: pr	otei	n ·		•						
		(x	i) s	EQUE	NCE :	DESC	RIPT	ION:	SEQ	ID :	NO:7	0:					
		Met 1	Asp	Val	Thr	Ser 5	Gln	Ala	Arg	Gly	Val	Gly	Leu	Glu	Met	Tyr 15	Pro
5		Gly	Thr	Ala	Gln 20	Pro	Ala	Ala	Pro	Asn 25	Thr	Thr	Ser	Pro	Glu 30	Leu	Asn
,	•	Leu	Ser	His 35	Pro	Leu	Leu	Gly	Thr 40	Ala	Leu	Ala	Asn	Gly 45	Thr	Gly	Glu
10		Leu	Ser 50	Glu	His	Gln	Gln	Tyr 55	Val	Ile	Gly	Leu	Phe 60	Leu	Ser	Cys	Leu
•		Tyr 65	Thr	Ile	Phe	Leu	Phe 70	Pro	Ile	Gly	Phe	Val 75	Gly	Asn	Ile	Leu	Ile 80
		Leu	Val	Val	Asn	Ile 85	Ser	Phe	Arg	Glu	Lys 90	Met	Thr	Ile	Pro	Asp 95	Leu
15		Tyr	Phe	Ile	Asn 100	Leu	Ala	Val	Ala	Asp 105	Leu	Ile	Leu	Val	Ala 110	Asp	Ser
		Leu	Ile	Glu 115	Val	Phe	Asn		His 120		Arg	Tyr	Tyr	Asp 125	Ile	Ala	Val
20	•	Leu	Cys 130	Thr	Phe	Met	Ser	Leu 135	Phe	Leu	Gln	Val	Asn 140	Met	Tyr	Ser	Ser
		Val 145	Phe	Phe	Leu	Thr	Trp 150	Met	Ser	Phe	Asp	Arg 155	Tyr	Ile	Ala	Leu	Ala 160
		Arg	Ala	Met	Arg	Cys 165	Ser	Leu	Phe	Arg	Thr 170	Lys	His	His	Ala	Arg 175	Leu
25		Ser	Cys	Gly	Leu 180	Ile	Trp	Met	Ala	Ser 185	Val	Ser	Ala	Thr	Leu 190	Val	Pro
	٠.	Phe	Thr.	Ala 195	Val	His	Leu	Gln	His 200	Thr	Asp	Glu	Ala	Cys 205	Phẹ	Cys	Phe
30	٠	Ala	Asp 210	Val	Arg	Glu	Val	Gln 215	Trp	Leu	Glu	Val	Thr 220	Leu	Gly	Phe	Ile
		Val 225	Pro	Phe	Ala	Ile	Ile 230	Gly	Leu	Cys	Tyr	Ser 235	Leu	Ile	Val	Arg	Val 240
		Leu	Val	Arg	Ala	His 245	Arg	His	Arg	Gly	Leu 250	Arg	Pro	Arg	Arg	Gln 255	Lys
35		Ala	Leu	Arg	Met 260	Ile	Leu	Ala	Val	Val 265	Leu	Val	Phe	Phe	Val 270	Cys	Trp

		Leu	Pro	G1u 275	Asn	Val	Phe	Ile	Ser 280	Val	His	Leu	Leu	Gln 285	Arg	Thr	Gln	
		Pro	Gly 290	Ala	Ala	Pro	Cys	Lys 295		Ser	Phe	Arg	His 300	Ala	His	Pro	Leu	
5		Thr 305	Gly	His	Ile	Val	Asn 310	Leu	Thr	Ala	Phe	Ser 315	Asn	Ser	Cys	Leu	Asn 320	
		Pro	Leu	Ile	Tyr	Ser 325	Phe	Leu	Gly	Glu	Thr 330		Arg	Asp	Lys	Leu 335	Arg	
10		Leu	Tyr	Ile	Glu 340	Gln	Lys	Thr	Asn	Leu 345	Pro	Ala	Leu	Asn	Arg 350	Phe	Cys	
		His	Ala	Ala 355	Leu	Lys	Ala	Val	Ile 360	Pro	Asp	Ser	Thr	Glu 365	Gln	Ser	Asp	
•		Val	Arg 370	Phe	Ser	Ser	Ala	Val 375					٠		* 4.*	. 17	5. <u>.</u>	
15	(72)	INFO	ORMAT	rion	FOR	SEQ	ID 1	NO:73	L:									
20	• . •	(ii	(B)	LEN TYI STI TOI	NGTH: PE: I RANDE POLOG	30 nucle EDNES EY:]	base eic a SS: s linea	e pai acid singl ar A (ge	irs .e enomi		10:71	·:						
	ACAGA	ATTC	CC TG	STGTO	GTTI	TAC	CGC	CCAG							<i>i</i>			30
	(73)	INFO	RMAT	NOI	FOR	SEQ	ID N	NO:72	: :				,					
25		(i)	(B) (C)	LEN TYI STR	CE CH NGTH: PE: IN RANDE POLOG	30 ucle	base ic a S: s	e pai acid singl	.rs			٠		,	, *			
30		(ii	.) MC	LECU	LE I	YPE:	DNA	A (ge	nomi	.c)					•	·		
		(xi	.) SE	QUEN	ICE D	ESCR	IPTI					: .* .		•				
	CTCGG	SATCO	A GG	CAGA	AGAG	TCG	CCT			÷ a				at '	•			30
	(74)	INFO	RMAT	NOI	FOR	SEQ	ID N	10:73	:									
35		(i)	(B)	LEN TYP	E CH IGTH: E: n	113 ucle	7 ba	se p	airs									

57

- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

٠.;	ATGGACCTGG	GGAAACCAAT	GAAAAGCGTG	CTGGTGGTGG	CTCTCCTTGT	CATTTTCCAG	60
5	GTATGCCTGT	GTCAAGATGA	GGTCACGGAC	GATTACATCG	GAGACAACAC	CACAGTGGAC	120
	TACACTTTGT	TCGAGTCTTT	GTGCTCCAAG	AAGGACGTGC	GGAACTTTAA	AGCCTGGTTC	180
	CTCCCTATCA	TGTACTCCAT	CATTTGTTTC	GTGGGCCTAC	TGGGCAATGG	GCTGGTCGTG	240
	TTGACCTATA	TCTATTTCAA	GAGGCTCAAG	ACCATGACCG	ATACCTACCT	GCTCAACCTG	300
	GCGGTGGCAG	ACATCCTCTT	CCTCCTGACC	CTTCCCTTCT	GGGCCTACAG	CGCGGCCAAG	360
10	TCCTGGGTCT	TCGGTGTCCA	CTTTTGCAAG	CTCATCTTTG	CCATCTACAA	GATGAGCTTC	420
	TTCAGTGGCA	TGCTCCTACT	TCTTTGCATC	AGCATTGACC	GCTACGTGGC	CATCGTCCAG	480
	GCTGTCTCAG	CTCACCGCCA	CCGTGCCCGC	GTCCTTCTCA	TCAGCAAGCT	GTCCTGTGTG	540
	GGCATCTGGA	TACTAGCCAC	AGTGCTCTCC	ATCCCAGAGC	TCCTGTACAG	TGACCTCCAG	600
	AGGAGCAGCA	GTGAGCAAGC	GATGCGATGC	TCTCTCATCA	CAGAGCATGT	GGAGGCCTTT	660
15	ATCACCATCC	AGGTGGCCCA	GATGGTGATC	GGCTTTCTGG	TCCCCCTGCT	GGCCATGAGC	720
	TTCTGTTACC	TTGTCATCAT	CCGCACCCTG	CTCCAGGCAC	GCAACTTTGA	GCGCAACAAG	780
	GCCATCAAGG	TGATCATCGC	TGTGGTCGTG	GTCTTCATAG	TCTTCCAGCT	GCCCTACAAT	840
	GGGGTGGTCC	TGGCCCAGAC	GGTGGCCAAC	TTCAACATCA	CCAGTAGCAC	CTGTGAGCTC	900
	AGTAAGCAAC	TCAACATCGC	CTACGACGTC	ACCTACAGCC	TGGCCTGCGT	CCGCTGCTGC	960
20	GTCAACCCTT	TCTTGTACGC	CTTCATCGGC	GTCAAGTTCC	GCAACGATCT	CTTCAAGCTC	1020
	TTCAAGGACC	TGGGCTGCCT	CAGCCAGGAG	CAGCTCCGGC	AGTGGTCTTC	CTGTCGGCAC	1080
	ATCCGGCGCT	CCTCCATGAG	TGTGGAGGCC	GAGACCACCA	CCACCTTCTC	CCCATAG	1137
						•	

(75) INFORMATION FOR SEQ ID NO:74:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 378 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- 30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

							. •		٠٠.						:		
		Met 1	Asp	Leu	Gly	Lys 5	Pro	Met	Lys	Ser	Val 10	Leu	Val	Val	Ala	Leu 15	Leu
		Val	Ile	Phe	Gln 20	Val	Cys	Leu	Cys	Cln 25	Asp	Glu	Val	Thr	Asp 30	Asp	Tyr
	5	Ile	Gly	Asp 35	Asn	Thr	Thr	Val	Asp 40	Tyr	Thr	Leu	Phe	Glu 45	Ser	Leu	Сув
		Ser	Lys 50	Lys	Asp	Val	Arg	Asn 55	Phe	Lys	Ala	Trp	Phe 60	Leu	Pro	Ile	Met
	10	Tyr 65	Ser	Ile	Ile	Cys	Phe 70	Val	Gly	Leu	Leu	Gly 75	Asn	Gly	Leu	Val	Val 80
		Leu	Thr	Tyr	Ile	Tyr 85	Phe	Lys	Arg	Leu	Lys 90	Thr	Met	Thr	Asp	Thr 95	Tyr
		Leu	Leu	Asn	Leu 100	Ala	Val	Ala	Asp	Ile 105	Leu	Phe	Leu	Leu	Thr 110	Leu	Pro
	15	Phe	Trp	Ala 115	Tyr	Ser	Ala	Ala	Lys 120	Ser	Trp	Val	Phe	Gly 125	Val	His	Phe
		Cys	Lys 130	Leu	Ile	Phe		Ile 135	Tyr	Lys	Met	Ser	Phe 140	Phe	Ser	Gly	Met
	20	Leu 145	Leu	Leu	Leu	Cys	Ile 150	Ser	Ile	Asp	Arg	Tyr 155	Val	Ala	Ile	Val	Gln 160
		Ala	Val	Ser	Ala	His 165	Arg	His	Arg	Ala	Arg 170	Val	Leu	Leu	Ile	Ser 175	Lys
42		Leu	Ser	Cys	Val 180	Gly	Ile	Trp	Ile	Leu 185	Ala	Thr	Val	Leu	Ser 190	Ile	Pro
	25	Glu	Leu	Leu 195	Tyr	Ser	Asp	Leu	Gln 200	Arg	Ser	Ser	Ser	Glu 205	Gln	Ala	Met
		Arg	Cys 210	Ser	Leu	Ile	Thr	Glu 215	His	Val	Glu	Ala	Phe 220	Ile	Thr	Ile	Gln
	30	Val 225	Ala	Gln	Met	Val	Ile 230	Gly	Phe	Leu	Val	Pro 235	Leu	Leu	Ala	Met	Ser 240
		Phe	Cys	Tyr	Leu	Val 245	Ile	Ile	Arg	Thr	Leu 250	Leu	Gln	Ala	Arg	Asn 255	Phe
			Arg		260					265					270		
	35	Ile	Val	Phe 275	Gln	Leu	Pro	Tyr	Asn 280	Gly	Val	Val	Leu	Ala 285	Gln	Thr	Val

		Ala	Asn 290	Phe	Asn	Ile	Thr	Ser 295	Ser	Thr	Cys	Glu	Leu 300	Ser	Lys	Gln	Leu	
	•	Asn 305	Ile	Ala	Tyr	Asp	Val 310	Thr	Tyr	Ser	Leu	Ala 315	Cys	Val	Arg	Cys	Cys 320	
5			Asn	Pro	Phe	Leu 325	Tyr		Phe		Gly 330	Val	Lys	Phe	Arg	Asn 335	Asp	
	•.	Leu	Phe	Lys	Leu 340	Phe	Lys	Asp	Leu	Gly 345	Cys	Leu	Ser	Gln	Glu 350		Leu	
10		Arg	Gln	Trp 355	Ser	Ser	Cys	Arg	His 360	Ile	Arg	Arg	Ser	Ser 365	Met	Ser	Val	
		Glu	Ala 370	Glu	Thr	Thr	Thr	Thr 375	Phe	Ser	Pro							
	(76)	INFO	RMAT	rion	FOR	SEQ	ID N	10:75	5:									
15		·(i)	_	-	CE CH													
15			(B)	TY	NGTH: PE: r	ucle	eic a	acid										. :
					POLOG			_	.e									
		(ii	.) MC	LECU	ле 1	YPE:	DNA	ı (ge	nomi	.c)								
20		(xi	.) SE	EQUE1	ICE I	ESCF	RIPTI	ON:	SEQ	ID N	10 : 75	5 :						ز. د
	CTGG	ATTC	A CC	TGG	ACCAC	CAC	CAAT	rgga	TA									32
	(77)	INFO	RMAI	NOI	FOR	SEQ	ID N	IO: 76	i :									
		(i)		-	CE CH													
25			(B)	TYE	E: r	ucle	eic a	cid										
					POLOC				.e									
		(ii	.) MC	LECU	LE 1	YPE:	DNA	4 (ge	nomi	.c)					٠			
		(xi) SE	QUEN	ICE I	ESCF	RIPTI	ON:	SEQ	ID N	10:76	i :						
30	CTCGG	ATCO	T GO	CAAAC	TTTC	TCF	TAC	AGTT										3 O
	(78)	INFO	RMAI	NOI	FOR	SEQ	ID 1	10:77	':									
35		(i)	(A) (B) (C)	LEN TYP STF	CE CH NGTH: PE: IN RANDE POLOG	108 nucle DNES	85 ba eic a 88: s	ase p acid singl	airs	3				·				
		(ii			JLE T				enomi	.c)								

The state of the s

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

	ATGGATATAC	AAATGGCAAA	CAATTTTACT	CCGCCCTCTG.	CAACTCCTCA	GGGAAATGAC	60
	TGTGACCTCT	ATGCACATCA	CAGCACGGCC	AGGATAGTAA		TTACAGCCTC	120
•	GTCTTCATCA	TTGGGCTCGT	GGGAAACTTA		•	TCAAAACAGG	180
5	ааааааатса	ACTCTACCAC	CCTCTATTCA	ACAAATTTGG	TGATTTCTGA	TATACTTTTT	240
	ACCACGGCTT	TGCCTACACG	AATAGCCTAC	TATGCAATGG	GCTTTGACTG	GAGAATCGGA	300
	GATGCCTTGT	GTAGGATAAC	TGCGCTAGTG	TTTTACATCA	ACACATATGC	AGGTGTGAAC	360
	TTTATGACCT	GCCTGAGTAT	TGACCGCTTC	ATTGCTGTGG	TGCACCCTCT	ACGCTACAAC	420
•.	AAGATAAAAA	GGATTGAACA	TGCAAAAGGC	GTGTGCATAT	TTGTCTGGAT	TCTAGTATTT	480
10	GCTCAGACAC	TCCCACTCCT	CATCAACCCT	ATGTCAAAGC	AGGAGGCTGA		540
	TGCATGGAGT	ATCCAAACTT	TGAAGAAACT	AAATCTCTTC	CCTGGATTCT	GCTTGGGGCA	600
.*	TGTTTCATAG	GATATGTACT	TCCACTTATA	ATCATTCTCA	TCTGCTATTC	TCAGATCTGC	660
	TGCAAACTCT	TCAGAACTGC	CAAACAAAAC	CCACTCACTG	AGAAATCTGG	TGTAAACAAA	720
	AAGGCTCTCA	ACACAATTAT	TCTTATTATT	GTTGTGTTTG	TTCTCTGTTT	CACACCTTAC	780
15	CATGTTGCAA	TTATTCAACA	TATGATTAAG	AAGCTTCGTT	TCTCTAATTT	CCTGGAATGT	840
	AGCCAAAGAC	ATTCGTTCCA	GATTTCTCTG	CACTTTACAG	TATGCCTGAT	GAACTTCAAT	900
	TGCTGCATGG	ACCCTTTTAT	CTACTTCTTT	GCATGTAAAG	GGTATAAGAG	AAAGGTTATG	960
	AGGATGCTGA	AACGGCAAGT	CAGTGTATCG	ATTTCTAGTG	CTGTGAAGTC	AGCCCCTGAA	1020
	GAAAATTCAC	GTGAAATGAC	AGAAACGCAG	ATGATGATAC	ATTCCAAGTC	TTCAAATGGA	1080
20	AAGTGA						1086

(79) INFORMATION FOR SEQ ID NO:78:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 361 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:
- Met Asp Ile Gln Met Ala Asn Asn Phe Thr Pro Pro Ser Ala Thr Pro 30 1 5 10 15

*#" *#"

	Gln	Gly	Asn	Asp 20	Cys	Asp	Leu	Tyr	Ala 25	His	His	Ser	Thr	Ala 30	Arg	Ile
	Val	Met	Pro 35	Leu	His	Tyr	Ser	Leu 40	Val	Phe	Ile	Ile	Gly 45	Leu	Val	Gly
. 5		Leu 50	Leu	Ala	Leu	Val	Val 55	Ile	Val	Gln	Asn	Arg 60	Lys	Lys	Ile	Asn
	Ser 65	Thr	Thr	Leu	Tyr	Ser 70	Thr	Asn	Leu	Val	Ile 75	Ser	Asp	Ile	Leu	Phe 80
10	Thr	Thr	Ala	Leu	Pro 85	Thr	Arg	Ile	Ala	Tyr 90	Tyr	Ala	Met	Gly	Phe 95	Asp
	Trp	Arg	Ile	Gly 100	Asp	Ala	Leu	Cys	Arg 105	Ile	Thr	Ala	Leu	Val 110	Phe	Tyr
	Ile	Asn	Thr 115	Tyr	Ala	Gly	Val	Asn 120	Phe	Met	Thr	Cys	Leu 125	Ser	Ile	Asp
15	Arg	Phe 130	Ile	Ala	Val	Val	His 135	Pro	Leu	Arg	Tyr	Asn 140	Lys	Ile	Lys	Arg
	Ile 145	Glu	His	Ala	Lys	Gly 150	Val	Cys	Ile	Phe	Val 155	Trp	Ile	Leu	Val	Phe 160
20	Ala	Gln	Thr	Leu	Pro 165	Leu	Leu	Ile	Asn	Pro 170	Met	Ser	Lys	Gln	Glu 175	Ala
	Glu	Arg	Ile	Thr 180	Cys	Met	Glu	Tyr	Pro 185	Asn	Phe	Glu	Glu	Thr 190	Lys	Ser
	Leu	Pro	Trp 195	Ile	Leu	Leu	Gly	Ala 200	Cys	Phe	Ile	Gly	Tyr 205	Val	Leu	Pro
25	Leu	Ile 210	Ile	Ile	Leu	Ile	Cys 215	Tyr	Ser	Gln	Ile	Cys 220	Cys	Lys	Leu	Phe
	Arg 225	Thr	Ala	Lys		Asn 230	Pro	Leu	Thr		Lys 235	Ser	Gly	Val	Asn	Lys 240
30	Lys	Ala	Leu	Asn	Thr 245	Ile	Ile	Leu	Ile	Ile 250	Val	Val	Phe	Val	Leu 255	Cys
	Phe	Thr	Pro	Tyr 260	His	Val	Ala	Ile	Ile 265	Gln	His	Met	Ile	Lys 270	Lys	Leu
	Arg	Phe	Ser 275	Asn	Phe	Leu	Glu	Cys 280	Ser	Gln	Arg	His	Ser 285	Phe	Gln	Ile
35	Ser	Leu 290	His	Phe	Thr	Val	Суs 295	Leu	Met	Asn	Phe	Asn 300	Cys	Cys	Met	Asp
	Pro	Phe	Ile	Tyr	Phe	Phe	Ala	Cys	Lys	Gly	Tyr	Lys	Arg	Lys	Val	Met

	305 310 315 320	
	Arg Met Leu Lys Arg Gln Val Ser Val Ser Ile Ser Ser Ala Val Lys 325 330 335	
5	Ser Ala Pro Glu Glu Asn Ser Arg Glu Met Thr Glu Thr Gln Met Met 340 345 350	
	Ile His Ser Lys Ser Ser Asn Gly Lys 355 360	
	(80) INFORMATION FOR SEQ ID NO:79:	
10	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
•	(ii) MOLECULE TYPE: DNA (genomic)	
15	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:	
	CTGGAATTCT CCTGCTCATC CAGCCATGCG G	31
	(81) INFORMATION FOR SEQ ID NO:80:	
20	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:	
25	CCTGGATCCC CACCCCTACT GGGGCCTCAG	30
	(82) INFORMATION FOR SEQ ID NO:81:	
30	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1446 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:	
	ATGCGGTGGC TGTGGCCCCT GGCTGTCTCT CTTGCTGTGA TTTTGGCTGT GGGGCTAAGC	60
35	AGGGTCTCTG GGGGTGCCCC CCTGCACCTG GGCAGGCACA GAGCCGAGAC CCAGGAGCAG	120

באוכטטטוט אוט מטטיטטעין י

63

	CAGAGCCGAT	CCAAGAGGGG	CACCGAGGAT	GAGGAGGCCA	AGGGCGTGCA	GCAGTATGTG	180
	CCTGAGGAGT	GGGCGGAGTA	CCCCCGGCCC	ATTCACCCTG	CTGGCCTGCA	GCCAACCAAG	240
	CCCTTGGTGG	CCACCAGCCC	TAACCCCGAC	AAGGATGGGG	GCACCCCAGA	CAGTGGGCAG	300
	GAACTGAGGG	GCAATCTGAC	AGGGGCACCA	GGGCAGAGGC	TACAGATCCA	GAACCCCCTG	360
5	TATCCGGTGA	CCGAGAGCTC	CTACAGTGCC	TATGCCATCA	TGCTTCTGGC	GCTGGTGGTG	420
	TTTGCGGTGG	GCATTGTGGG	CAACCTGTCG	GTCATGTGCA	TCGTGTGGCA	CAGCTACTAC	480
	CTGAAGAGCG	CCTGGAACTC	CATCCTTGCC	AGCCTGGCCC	TCTGGGATTT	TCTGGTCCTC	540
	TTTTTCTGCC	TCCCTATTGT	CATCTTCAAC	GAGATCACCA	AGCAGAGGCT	ACTGGGTGAC	600
	GTTTCTTGTC	GTGCCGTGCC	CTTCATGGAG	GTCTCCTCTC	TGGGAGTCAC	GACTTTCAGC	660
10	CTCTGTGCCC	TGGGCATTGA	CCGCTTCCAC	GTGGCCACCA	GCACCCTGCC	CAAGGTGAGG	720
	CCCATCGAGC	GGTGCCAATC	CATCCTGGCC	AAGTTGGCTG	TCATCTGGGT	GGGCTCCATG	780
	ACGCTGGCTG	TGCCTGAGCT	CCTGCTGTGG	CAGCTGGCAC	AGGAGCCTGC	CCCCACCATG	840
	GGCACCCTGG	ACTCATGCAT	CATGAAACCC	TCAGCCAGCC	TGCCCGAGTC	CCTGTATTCA	900
	CTGGTGATGA	CCTACCAGAA	CGCCCGCATG	TGGTGGTACT	TTGGCTGCTA	CTTCTGCCTG	960
15	CCCATCCTCT	TCACAGTCAC	CTGCCAGCTG	GTGACATGGC	GGGTGCGAGG	CCCTCCAGGG	1020
	AGGAAGTCAG	AGTGCAGGGC	CAGCAAGCAC	GAGCAGTGTG	AGAGCCAGCT	CAACAGCACC	1080
	GTGGTGGGCC	TGACCGTGGT	CTACGCCTTC	TGCACCCTCC	CAGAGAACGT	CTGCAACATC	1140
	GTGGTGGCCT	ACCTCTCCAC	CGAGCTGACC	CGCCAGACCC	TGGACCTCCT	GGGCCTCATC	1200
	AACCAGTTCT	CCACCTTCTT	CAAGGGCGCC	ATCACCCCAG	TGCTGCTCCT	TTGCATCTGC	1260
20	AGGCCGCTGG	GCCAGGCCTT	CCTGGACTGC	TGCTGCTGCT	GCTGCTGTGA	GGAGTGCGGC	1320
	GGGGCTTCGG	AGGCCTCTGC	TGCCAATGGG	TCGGACAACA	AGCTCAAGAC	CGAGGTGTCC	1380
	TCTTCCATCT	ACTTCCACAA	GCCCAGGGAG	TCACCCCCAC	TCCTGCCCCT	GGGCACACCT	1440
	TGCTGA						1446

(83) INFORMATION FOR SEQ ID NO:82:

25 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 481 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant

30 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

		Met 1	Arg	Trp	Leu	Trp 5	Pro	Leu	Ala	Val	Ser 10	Leu	Ala	Val	Ile	Leu 15	Ala
	5	Val	Gly	Leu	Ser 20	Arg	Val	Ser	Gly	Gly 25	Ala	Pro	Leu	His	Leu 30	Gly	Arg
	•	His	Arg	Ala 35	Glu	Thr	Gln	Glu	Gln 40	Gln	Ser	Arg	Ser	Lys 45	Arg	Gly	Thr
		Glu	Asp 50	Glu	Glu	Ala	Lys	Gly 55	Val	Gln	Gln	Tyr	Val 60	Pro	Glu	Glu	Trp
	10	Ala 65	Glu	Tyr	Pro	Arg	Pro 70	Ile	His	Pro	Ala	Gly 75	Leu	Gln	Pro	Thr	Lys 80
		Pro	Leu	Val	Ala	Thr 85	Ser	Pro	Asn	Pro	Asp 90	Lys	Asp	Gly	Gly	Thr 95	Pro
4.	15	Asp	Ser	Gly	Gln 100	Glu	Leu	Arg	Gly	Asn 105	Leu	Thr	Gly	Ala	Pro 110	Gly	Gln
÷		Arg	Leu	Gln 115	Ile	Gln	Asn	Pro	Leu 120	Tyr	Pro	Val	Thr	Glu 125	Ser	Ser	Tyr
.* · ·		Ser	Ala 130	Tyr	Ala	Ile	Met	Leu 135	Leu	Ala	Leu	Val	Val 140	Phe	Ala	Val	Gly
	20	Ile 145	Val	Gly	Asn	Leu	Ser 150	Val	Met	Cys	Ile	Val 155	Trp	His	Ser	Tyr	Tyr 160
:		Leu	Lys	Ser	Ala	Trp 165	Asn	Ser	Ile	Leu	Ala 170	Ser	Leu	Ala	Leu	Trp 175	Asp
	25	Phe	Leu	Val	Leu 180	Phe	Phe	Cys	Leu	Pro 185	Ile	Val	Ile	Phe	Asn 190	Glu	Ile
٠		Thr	Lys	Gln 195	Arg	Leu	Leu	Gly	Asp 200	Val	Ser	Cys	Arg	Ala 205	Val	Pro	Phe
		Met	Glu 210	Val	Ser	Ser	Leu	Gly 215	Val	Thr	Thr	Phe	Ser 220	Leu	Cys	Ala	Leu
	30	Gly 225	Ile	Asp	Arg	Phe	His 230	Val	Ala	Thr	Ser	Thr 235	Leu	Pro	Lys	Val	Arg 240
		Pro	Ile	Glu	Arg	Cys 245	Gln	Ser	Ile	Leu	Ala 250	Lys	Leu	Ala	Val	Ile 255	Trp
	35	Val	Gly	Ser	Met 260	Thr	Leu	Ala	Val	Pro 265	Glu	Leu	Leu	Leu	Trp 270	Gln	Leu
		Ala	Gln	Glu	Pro	Ala	Pro	Thr	Met	Gly	Thr	Leu	Asp	Ser	Cys	Ile	Met

				275					280					285			
		Lys	Pro 290	Ser	Ala	Ser	Leu	Pro 295	Glu	Ser	Leu	Tyr	Ser 300	Leu	Val	Met	Thr
5		Tyr 305		Asn	Ala	Arg	Met 310	Trp	Trp		Phe	Gly 315	Cys	Tyr	Phe	Cys	Leu 320
		Pro	Ile	Leu	Phe	Thr 325	Val	Thr	Cys	Gln	Leu 330	Val	Thr	Trp	Arg	Val 335	Arg
	•	Gly	Pro	Pro	Gly 340	Arg	Lys	Ser	Glu	Cys 345	Arg	Ala	Ser	Lys	His 350	Glu	Gln
10		Cys	Glu	Ser 355	Gln	Leu	Asn	Ser	Thr 360	Val	Val	Gly	Leu	Thr 365	Val	Val	Tyr
· ;		Ala	Phe 370	-	Thr	Leu	Pro	Glu 375	Asn	Val	Cys		Ile 380	Val	Val	Ala	Tyr
15		Leu 385	Ser	Thr	Glu	Leu	Thr 390	Arg	Gln	Thr	Leu	Asp 395	Leu	Leu	Gly	Leu	Ile 400
		Asn	Gln	Phe	Ser	Thr 405	Phe	Phe	Lys	Gly	Ala 410	Ile	Thr	Pro	Val	Leu 415	Leu
	•	Leu	Cys	Ile	Cys 420		Pro	Leu	Gly	Gln 425		Phe	Leu	Asp	Cys 430	Cys	Cys
20		Cys	Cys	Cys 435		Glu	Glu	Cys	Gly 440		Ala	Ser	Glu	Ala 445	Ser	Ala	Ala
,		Asn	Gly 450		Asp	Asn	Lys	Leu 455		Thr	Glu	Val	Ser 460		Ser	Ile	Tyr
25		Phe		Lys	Pro	Arg	Glu 470		Pro	Pro	Leu	Leu 475	Pro	Leu	Gly	Thr	Pro 480
		Cys											-				
1	(84)	INF	ORMA	MOITA	FOR	SEC) ID	ио:8	3:								
30		(i	(E (C	A) LE 3) TY C) ST	NGTH PE: RANI	nucl	CTER bas eic ESS: line	e pa ació sing	irs	•							
		()	.i) N	OLEC	TULE	TYPE	E: D1	g) Al	renon	nic)							
35		(3	(i) 9	SEOUE	ENCE	DESC	CRIPT	OION:	SEC	D ID	NO:8	3:					

	(85) INFORMATION FOR SEQ ID NO:84:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 22 base pairs	
	(B) TYPE: nucleic acid	
5	(C) STRANDEDNESS: single	
,		
	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:	
	TCATGTATTA ATACTAGATT CT	22
10	(86) INFORMATION FOR SEQ ID NO:85:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 38 base pairs	
	(B) TYPE: nucleic acid	, •
	(C) STRANDEDNESS: single	
15	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:	
•	(AI) DEGOENCE PERCENTITION. DEG ID NO.03.	
	TACCATGTGG AACGCGACGC CCAGCGAAGA GCCGGGGT	38
	(87) INFORMATION FOR SEQ ID NO:86:	
20	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs	
	(B) TYPE: nucleic acid (C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
•	The state of the s	
25	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:	
	CGGAATTCAT GTATTAATAC TAGATTCTGT CCAGGCCCG	39
	(88) INFORMATION FOR SEQ ID NO:87:	
	(i) SEQUENCE CHARACTERISTICS:	
30	(A) LENGTH: 1101 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	•
35	(vi) SECTENCE DESCRIPTION: SEC ID NO.87.	

67

		ATGTGGAACG	CGACGCCCAG	CGAAGAGCCG	GGGTTCAACC	TCACACTGGC	CGACCTGGAC	60
		TGGGATGCTT	CCCCCGGCAA	CGACTCGCTG	GGCGACGAGC	TGCTGCAGCT	CTTCCCCGCG	120
		CCGCTGCTGG	CGGGCGTCAC	AGCCACCTGC	GTGGCACTCT	TCGTGGTGGG	TATCGCTGGC	180
		AACCTGCTCA	CCATGCTGGT	GGTGTCGCGC	TTCCGCGAGC	TGCGCACCAC	CACCAACCTC	240
	5	TACCTGTCCA	GCATGGCCTT	CTCCGATCTG	CTCATCTTCC	TCTGCATGCC	CCTGGACCTC	300
		GTTCGCCTCT	GGCAGTACCG	GCCCTGGAAC	TTCGGCGACC	TCCTCTGCAA	ACTCTTCCAA	360
		TTCGTCAGTG	AGAGCTGCAC	CTACGCCACG	GTGCTCACCA	TCACAGCGCT	GAGCGTCGAG	420
		CGCTACTTCG	CCATCTGCTT	CCCACTCCGG	GCCAAGGTGG	TGGTCACCAA	GGGCGGGTG	480
		AAGCTGGTCA	TCTTCGTCAT	CTGGGCCGTG	GCCTTCTGCA	GCGCCGGGCC	CATCTTCGTG	540
	10	CTAGTCGGGG	TGGAGCACGA	GAACGGCACC	GACCCTTGGG	ACACCAACGA	GTGCCGCCCC	600
		ACCGAGTTTG	CGGTGCGCTC	TGGACTGCTC	ACGGTCATGG	TGTGGGTGTC	CAGCATCTTC	660
		TTCTTCCTTC	CTGTCTTCTG	TCTCACGGTC	CTCTACAGTC	TCATCGGCAG	GAAGCTGTGG	720
		CGGAGGAGGC	GCGGCGATGC	TGTCGTGGGT	GCCTCGCTCA	GGGACCAGAA	CCACAAGCAA	780
. :		ACCGTGAAAA	TGCTGGCTGT	AGTGGTGTTT	GCCTTCATCC	TCTGCTGGCT	CCCCTTCCAC	840
	15	GTAGGGCGAT	ATTTATTTTC	CAAATCCTTT	GAGCCTGGCT	CCTTGGAGAT	TGCTCAGATC	900
		AGCCAGTACT	GCAACCTCGT	GTCCTTTGTC	CTCTTCTACC	TCAGTGCTGC	CATCAACCCC	960
		ATTCTGTACA	ACATCATGTC	CAAGAAGTAC	CGGGTGGCAG	TGTTCAGACT	TCTGGGATTC	1020
		GAACCCTTCT	CCCAGAGAAA	GCTCTCCACT	CTGAAAGATG	AAAGTTCTCG	GGCCTGGACA	1080
		GAATCTAGTA	TTAATACATG	A				1101

20 (89) INFORMATION FOR SEQ ID NO:88:

25

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 366 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

Met Trp Asn Ala Thr Pro Ser Glu Glu Pro Gly Phe Asn Leu Thr Leu 1 5 10 15

Ala Asp Leu Asp Trp Asp Ala Ser Pro Gly Asn Asp Ser Leu Gly Asp 20 25 30

	.•	Glu	Leu	Leu 35	Gin	Leu	Pne	Pro	A1a 40	Pro:	Leu	ьeu	Ala	45	vaı	Inr	Ala
		Thr	Cys 50	Val	Ala	Leu	Phe	Val 55	Val	Gly	Ile	Ala	Gly 60	Asn	Leu	Leu	Thr
5		Met 65	Leu	Val	Val	Ser	Arg 70	Phe	Arg	Glu	Leu	Arg 75	Thr	Thr	Thr	Asn	Lev 80
		Tyr	Leu	Ser	Ser	Met 85	Ala	Phe	Ser	Asp	Leu 90	Leu	Ile	Phe	Leu	Cys 95	Met
10		Pro	Leu	Asp	Leu 100	Val	Arg	Leu ·	Trp	Gln 105	Tyr	Arg	Pro	Trp	Asn 110	Phe	Gly
		Asp	Leu	Leu 115	Cys	Lys	Leu	Phe	Gln 120	Phe	Val	Ser	Glu	Ser 125	Cys	Thr	Туз
		Ala	Thr 130	Val	Leu	Thr	Ile	Thr 135	Ala	Leu	Ser	Val	Glu 140	Arg	Tyr	Phe	Ala
15		Ile 145	Cys	Phe	Pro	Leu	Arg 150	Ala	Lys	Val	Val	Val 155	Thr	Lys	Gly	Arg	Va]
		Lys	Leu	Val	Ile	Phe 165	Val	Ile	Trp	Ala	Val 170	Ala	Phe	Cys	Ser	Ala 175	Gly
20		Pro	Ile	Phe	Val 180	Leu	Val	Gly	Val	Glu 185	His	Glu	Asn	Gly	Thr 190	Asp	Pro
		Trp	Asp	Thr 195	Asn	Glu	Cys	Arg	Pro 200	Thr	Glu	Phe	Ala	Val 205	Arg	Ser	Gl
		Leu	Leu 210		Val	Met	Val	Trp 215	Val	Ser	Ser	Ile	Phe 220		Phe	Leu	Pro
25		Val 225		Cys	Leu	Thr	Val 230	Leu	Tyr	Ser	Leu	Ile 235	Gly	Arg	Lys	Leu	Tr _]
		Arg	Arg	Arg	Arg	Gly 245		Ala	Val	Val	Gly 250	Ala	Ser	Leu	Arg	Asp 255	
30		Asń	His	Lys	Gln 260		Val	Lys	Met	Leu 265		Val	Val	Val	Phe 270	Ala	Ph
		Ile	Leu	Cys 275	_	Leu	Pro	Phe	His 280		Gly	, Arg	Tyr	Leu 285		Ser	Ly
			290)				295					300				
35		305	i				310	•	Phe			315					32
		Ile	Lev	TVY	Asn	lle	Met	Ser	Lvs	Lvs	Tyr	Arq	val	Ala	val	Pne	Ar

PCT/US99/23938

	335	
	Leu Leu Gly Phe Glu Pro Phe Ser Gln Arg Lys Leu Ser Thr Leu Ly 340 345 350	s
5	Asp Glu Ser Ser Arg Ala Trp Thr Glu Ser Ser Ile Asn Thr 355 360 365	
	(90) INFORMATION FOR SEQ ID NO:89:	
10	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:	
	GCAAGCTTGT GCCCTCACCA AGCCATGCGA GCC	33
15	(91) INFORMATION FOR SEQ ID NO:90:	,
20	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:	
	CGGAATTCAG CAATGAGTTC CGACAGAAGC	30
	(92) INFORMATION FOR SEQ ID NO:91:	
25	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1842 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
30	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:	
	ATGCGAGCCC CGGGCGCGCT TCTCGCCCGC ATGTCGCGGC TACTGCTTCT GCTACTGCTC	60
	AAGGTGTCTG CCTCTGCGGGGTC GCCCCTGCGT CCAGAAACGA AACTTGTCTG	120
	GGGGAGAGCT GTGCACCTAC AGTGATCCAG CGCCGCGGCA GGGACGCCTG GGGACCGGGA	180
35	AATTCTGCAA GAGACGTTCT GCGAGCCCGA GCACCCAGGG AGGAGCAGGG GGCAGCCTTT	240

CTTGCGGGAC CCTCCTGGGA CCTGCCGGCG GCCCCGGGCC GTGACCCGGC TGCAGGCAGA 300 GGGGCGGAGG CGTCGGCAGC CGGACCCCCG GGACCTCCAA CCAGGCCACC TGGCCCCTGG 360 AGGTGGAAAG GTGCTCGGGG TCAGGAGCCT TCTGAAACTT TGGGGAGAGG GAACCCCACG 420 GCCCTCCAGC TCTTCCTTCA GATCTCAGAG GAGGAAGAGA AGGGTCCCAG AGGCGCTGGC ATTTCCGGGC GTAGCCAGGA GCAGAGTGTG AAGACAGTCC CCGGAGCCAG CGATCTTTTT TACTGGCCAA GGAGAGCCGG GAAACTCCAG GGTTCCCACC ACAAGCCCCT GTCCAAGACG 600 GCCAATGGAC TGGCGGGGCA CGAAGGGTGG ACAATTGCAC TCCCGGGCCG GGCGCTGGCC CAGAATGGAT CCTTGGGTGA AGGAATCCAT GAGCCTGGGG GTCCCCGCCG GGGAAACAGC 720 ACGAACCGGC GTGTGAGACT GAAGAACCCC TTCTACCCGC TGACCCAGGA GTCCTATGGA 10 GCCTACGCGG TCATGTGTCT GTCCGTGGTG ATCTTCGGGA CCGGCATCAT TGGCAACCTG 840 GCGGTGATGA GCATCGTGTG CCACAACTAC TACATGCGGA GCATCTCCAA CTCCCTCTTG GCCAACCTGG CCTTCTGGGA CTTTCTCATC ATCTTCTTCT GCCTTCCGCT GGTCATCTTC 960 CACGAGCTGA CCAAGAAGTG GCTGCTGGAG GACTTCTCCT GCAAGATCGT GCCCTATATA 1020 GAGGTCGCTT CTCTGGGAGT CACCACTTTC ACCTTATGTG CTCTGTGCAT AGACCGCTTC 1080 15 CGTGCTGCCA CCAACGTACA GATGTACTAC GAAATGATCG AAAACTGTTC CTCAACAACT 1140 GCCAAACTTG CTGTTATATG GGTGGGAGCT CTATTGTTAG CACTTCCAGA AGTTGTTCTC 1200 CGCCAGCTGA GCAAGGAGGA TTTGGGGTTT AGTGGCCGAG CTCCGGCAGA AAGGTGCATT 1260 ATTAAGATCT CTCCTGATTT ACCAGACACC ATCTATGTTC TAGCCCTCAC CTACGACAGT 1320 GCGAGACTGT GGTGGTATTT TGGCTGTTAC TTTTGTTTGC CCACGCTTTT CACCATCACC 1380 20 TGCTCTCTAG TGACTGCGAG GAAAATCCGC AAAGCAGAGA AAGCCTGTAC CCGAGGGAAT 1440 AAACGGCAGA TTCAACTAGA GAGTCAGATG AACTGTACAG TAGTGGCACT GACCATTTTA 1500 TATGGATTTT GCATTATTCC TGAAAATATC TGCAACATTG TTACTGCCTA CATGGCTACA 1560 GGGGTTTCAC AGCAGACAAT GGACCTCCTT AATATCATCA GCCAGTTCCT TTTGTTCTTT 1620 AAGTCCTGTG TCACCCCAGT CCTCCTTTTC TGTCTCTGCA AACCCTTCAG TCGGGCCTTC 1680 25 ATGGAGTGCT GCTGCTGTTG CTGTGAGGAA TGCATTCAGA AGTCTTCAAC GGTGACCAGT 1740 GATGACAATG ACAACGAGTA CACCACGGAA CTCGAACTCT CGCCTTTCAG TACCATACGC 1800 CGTGAAATGT CCACTTTTGC TTCTGTCGGA ACTCATTGCT GA 1842

(93) INFORMATION FOR SEQ ID NO:92:

5	· (i	(A (B (C	QUEN LE TY ST TO	NGTH PE: RAND	: 61 amin EDNE	3 am o ac SS:	ino id	acid								
	(i	.i) M	OLEC	OLE '	TYPE	: pr	otei	n					•			
	(x	i) S	EQUE:	NCE :	DESC	RIPT	ION:	SEQ	ID :	NO : 9	2:					
	Met 1	Arg	Ala	Pro	Gly 5	Ala	Leu	Leu	Ala	Arg 10	Met	Ser	Arg	Leu	Leu 15	Let
10	Leu	Leu	Leu	Leu 20	Lys	Val	Ser	Ala	Ser 25	Ser	Ala	Leu	Gly	Val 30	Ala	Pro
	Ala	Ser	Arg 35	Asn	Glu	Thr	Cys	Leu 40	Gly	Glu	Ser	Cys	Ala 45	Pro	Thr	Val
15	Ile	Gln 50	Arg	Arg	Gly	Arg	Asp 55	Ala	Trp	Gly	Pro	Gly 60	Asn	Ser	Ala	Arg
	Asp 65	Val	Leu	Arg	Ala	Arg 70	Ala	Pro	Arg	Glu	Glu 75	Gln	Gly	Ala	Ala	Phe 80
	Leu	Ala	Gly	Pro	Ser 85	Trp	Asp	Leu	Pro	Ala 90	Ala	Pro	Gly	Arg	Asp 95	Pro
20	Ala	Ala	Gly	Arg 100	Gly	Ala	Glu	Aļla	Ser 105	Ala	Ala	Gly	Pro	Pro 110	Gly	Pro
	Pro	Thr	Arg 115	Pro	Pro	Gly	Pro	Trp 120	Arg	Trp	Lys	Gly	Ala 125	Arg	Gly	Gln
25	Glu	Pro 130	Ser	Glu	Thr	Leu	Gly 135	Arg	Gly	Asn	Pro	Thr 140	Ala	Leu	Gln	Leu
	Phe 145	Leu	Gln	Ile	Ser	Glu 150	Glu	Glu	Glu	Lys	Gly 155	Pro	Arg	Gly	Ala	Gly 160
	Ile	Ser	Gly	Arg	Ser 165	Gln	Glu	Gln	Ser	Val 170	Lys	Thr	Val	Pro	Gly 175	Ala
30	Ser	Asp	Leu	Phe 180	Tyr	Trp	Pro	Arg	Arg 185	Ala	Gly	Lys	Leu	Gln 190	Gly	Ser
	His	His	Lys 195	Pro	Leu	Ser	Lys	Thr 200	Ala	Asn	Gly	Leu	Ala 205	Gly	His	Glu
35 ⁻	Gly	Trp 210	Thr	Ile	Ala	Leu	Pro 215	Gly	Arg	Ala	Leu	Ala 220	Gln	Asn	Gly	Ser
	Leu 225	Gly	Glu	Gly	Ile	His 230	Glu	Pro	Gly	Gly	Pro 235	Arg	Arg	Gly	Asn	Ser 240

									7	' 2							
		Thr	Asn	Arg	Arg	Val 245	Arg	Leu	Lys	Asn	Pro 250	Phe	Tyr	Pro	Leu	Thr 255	Gln
		Glu	Ser	Tyr	Gly 260	Ala	Tyr	Ala	Val	Met 265	Cys	Leu	Ser	Val	Val 270	Ile	Phe
	5 .	Gly	Thr	Gly 275	Ile	Ile	Gly	Asn	Leu 280	Ala	Val	Met	Ser	Ile 285	Val	Cys	His
		Asn	Tyr 290	Tyr	Met	Arg	Ser	Ile 295	Ser	Asn	Ser	Leu	Leu 300	Ala	Asn	Leu	Ala
	10	Phe 305	Trp	Asp	Phe	Leu	Ile 310	Ile	Phe	Phe	Cys	Leu 315	Pro	Leu	Val	Ile	Phe 320
		His	Glu	Leu	Thr	Lys 325	Lys	Trp	Leu	Leu	Glu 330	Asp	Phe	Ser	Cys	Lys 335	Ile
		Val	Pro	Tyr	Ile 340	Glu	Val	Ala	Ser	Leu 345	Gly	Val	Thr	Thr	Phe 350	Thr	Leu
	15	Cys	Ala	Leu 355	Cys	Ile	Asp	Arg	Phe 360	Arg	Ala	Ala	Thr	Asn 365	Val	Gln	Met
;		Tyr	Tyr 370	Glu	Met	Ile	Glu	Asn 375	Cys	Ser	Ser	Thr	Thr 380	Ala	Lys	Leu	Ala
	20	Val 385	Ile	Trp	Val	Gly	Ala 390	Leu	Leu	Leu	Ala	Leu 395	Pro	Glu	Val	Val	Leu 400
		Arg	Gln	Leu	Ser	Lys 405	Glu	Asp	Leu	Gly	Phe 410	Ser	Gly	Arg	Ala	Pro 415	Ala
		Glu	Arg	Cys	Ile 420	Ile	Lys	Ile	Ser	Pro 425	Asp	Leu	Pro	Asp	Thr 430	Ile	Tyr
	25	Val	Leu	Ala 435	Leu	Thr	Tyr	Asp	Ser 440	Ala	Arg	Leu	Trp	Trp 445	Tyr	Phe	Gly
		Cys	Tyr 450	Phe	Cys	Leu	Pro	Thr 455	Leu	Phe	Thr	Ile	Thr 460	Cys	Ser	Leu	Val
	30	Thr 465	Ala	Arg	Lys	Ile	Arg 470	Lys	Ala	Glu		Ala 475	Cys	Thr	Arg	Gly	Asn 480
		Lys	Arg	Gln	Ile					Gln 4			•	Thr		_	
		Leu	Thr	Ile	Leu 500	Tyr	Gly	Phe	Cys	Ile 505	Ile	Pro	Glu	Asn	Ile 510	Cys	Asn
	35	Ile	Val	Thr 515	Ala	Tyr	Met	Ala	Thr 520	Gly	Val	Ser	Gln	Gln 525	Thr	Met	Asp
		Leu	Leu	Asn	Ile	Ile	Ser	${\tt Gln}$	Phe	Leu	Leu	Phe	Phe	Lys	Ser	Cys	Val

	•	530		. ' . •		•	535		• ,			540	•				
	Th: 549	Pro	Val	Leu	Leu	Phe 550	Cys	Leu	Cys	Lys	Pro 555	Phe	Ser	Arg	Ala	Phe 560	
5	Met	Glu	Cys	Cys	Cys 565	Cys	Cys	Cys	Glu	Glu 570	Cys	Ile	Gln	Lys	Ser 575	Ser	
	Thi	· Val	Thr	Ser 580	Asp	Asp	Asn	Asp	Asn 585	Glu	Tyr	Thr	Thr	Glu 590	Leu	Glu	
	Let	Ser	Pro 595	Phe	Ser	Thr	Ile	Arg 600		Glu	Met	Ser	Thr 605		Ala	Ser	
10	Va]	Gly 610	Thr	His	Cys												
	(94) INF	ORMAT	CION	FOR	SEQ	ID N	10:93	3:									
	· (i	.) SEÇ	UENC	CE CH	IARAC	TER	STIC	S:									
15						base eic a	-	rs									و ً
•		(C)	STR	RANDE	EDNES	SS: S	ingl	.e									
		(D)	TOF	POLOG	3Y:]	inea	r										٠.
	(i	.i) MO	LECU	λΕ 1	YPE:	DNA	(ge	nomi	ic)						-		
	(2	:i) SE	QUEN	ICE I	ESCF	RIPTI	ON:	SEQ	ID N	10:93	3:						Ý
20	CAGAATTO	AG AG	AAAA	AAAC	G TGA	rata <i>l</i>	GGT	TTT	r								34
	(95) INF	ORMAT	ION	FOR	SEQ	ID N	IO: 94	·	٠	•	٠						
	(i) SEQ	-														
						base ic a	_	.rs									
25						SS: s		.e									-
		(D)	TOF	OLOG	Y: 1	inea	ır										
	(i	.i) MO	LECU	TE 1	YPE:	DNA	(ge	enomi	ic)								
	(xi) S	EQUE	ENCE	DESC	RIPI	: NOI	SEÇ	Q ID	NO: 9	4:						
	TTGGATCC	CT GG	TGCA	MAAC	CAA :	TGA	AGA	ΑT									32
30	(96) INF	ORMAT	NOI	FOR	SEQ	ID N	10 : 95	5:									
	(i) SEQ							_								
						8 ba		oairs	5								
٥.		(C)	STR	ZANDE	EDNES	SS: S	ingl	.e									
35		(D)	TOF	POLOG	Y: 1	inea	ır										
	(i	.i) MO	LECU	ILE I	TYPE:	DNA	(ge	mon	ic)								

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

		ATGGTTTTTG	CTCACAGAAT	GGATAACAGC	AAGCCACATT	TGATTATTCC	TACACTTCTG	60
		GTGCCCCTCC	AAAACCGCAG	CTGCACTGAA	ACAGCCACAC	CTCTGCCAAG	CCAATACCTG	120
		ATGGAATTAA	GTGAGGAGCA	CAGTTGGATG	AGCAACCAAA	CAGACCTTCA	CTATGTGCTG	180
	5	AAACCCGGGG	AAGTGGCCAC	AGCCAGCATC	TTCTTTGGGA	TTCTGTGGTT	GTTTTCTATC	240
		TTCGGCAATT	CCCTGGTTTG	TTTGGTCATC	CATAGGAGTA	GGAGGACTCA	GTCTACCACC	300
		AACTACTTTG	TGGTCTCCAT	GGCATGTGCT	GACCTTCTCA	TCAGCGTTGC	CAGCACGCCT	360
		TTCGTCCTGC	TCCAGTTCAC	CACTGGAAGG	TGGACGCTGG	GTAGTGCAAC	GTGCAAGGTT	420
		GTGCGATATT	TTCAATATCT	CACTCCAGGT	GTCCAGATCT	ACGTTCTCCT	CTCCATCTGC	480
	10	ATAGACCGGT	TCTACACCAT	CGTCTATCCT	CTGAGCTTCA	AGGTGTCCAG	AGAAAAAGCC	540
		AAGAAAATGA	TTGCGGCATC	GTGGATCTTT	GATGCAGGCT	TTGTGACCCC	TGTGCTCTTT	600
		TTCTATGGCT	CCAACTGGGA	CAGTCATTGT	AACTATTTCC	TCCCCTCCTC	TTGGGAAGGC	660
•		ACTGCCTACA	CTGTCATCCA	CTTCTTGGTG	GGCTTTGTGA	TTCCATCTGT	CCTCATAATT	720
	•	TTATTTTACC	AAAAGGTCAT	AAAATATATT	TGGAGAATAG	GCACAGATGG	CCGAACGGTG	780
	15	AGGAGGACAA	TGAACATTGT	CCCTCGGACA	AAAGTGAAAA	CTATCAAGAT	GTTCCTCATT	840
		TTAAATCTGT	TGTTTTTGCT	CTCCTGGCTG	CCTTTTCATG	TAGCTCAGCT	ATGGCACCCC	900
		CATGAACAAG	ACTATAAGAA	AAGTTCCCTT	GTTTTCACAG	CTATCACATG	GATATCCTTT	960
		AGTTCTTCAG	CCTCTAAACC	TACTCTGTAT	TCAATTTATA	ATGCCAATTT	TCGGAGAGGG	1020
		ATGAAAGAGA	CTTTTTGCAT	GTCCTCTATG	AAATGTTACC	GAAGCAATGC	CTATACTATC	1080
	20	ACAACAAGTT	CAAGGATGGC	CAAAAAAAAC	TACGTTGGCA	TTTCAGAAAT	CCCTTCCATG	1140
		GCCAAAACTA	TTACCAAAGA	CTCGATCTAT	GACTCATTTG	ACAGAGAAGC	CAAGGAAAAA	1200
		AAGCTTGCTT	GGCCCATTAA	CTCAAATCCA	CCAAATACTT	TTGTCTAA		1248

(97) INFORMATION FOR SEQ ID NO:96:

(i) SEQUENCE CHARACTERISTICS: 25

(A) LENGTH: 415 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- 30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

	Met 1	Val	Phe	Ala	His 5	Arg	Met	Asp	Asn	Ser 10	Lys	Pro	His	Leu	Ile 15	Ile
	Pro	Thr	Leu	Leu 20	Val	Pro	Leu	Gln	Asn 25	Arg	Ser	Cys	Thr	Glu 30	Thr	Ala
5	Thr	Pro '	Leu 35	Pro	Ser	Gln	Tyr	Leu 40	Met	Glu	Leu	Ser	Glu 45	Glu	His	Ser
	Trp	Met 50	Ser	Asn	Gln	Thr	Asp 55	Leu	His	Tyr	Val	Leu 60	Lys	Pro	Gly	Glu
10	Val 65	Ala	Thr	Ala	Ser	Ile 70	Phe	Phe	Gly	Ile	Leu 75	Trp	Leu	Phe	Ser	Ile 80
	Phe	Gly	Asn	Ser	Leu 85	Val	Cys	Leu	Val	Ile 90	His	Arg	Ser	Arg	Arg 95	Thr
·	Gln	Ser	Thr	Thr 100	Asn	Tyr	Phe	Val	Val 105	Ser	Met	Ala	Cys	Ala 110	Asp	Leu
15	Leu	Ile	Ser 115	Val	Ala	Ser	Thr	Pro 120	Phe	Val	Leu	Leu	Gln 125	Phe	Thr	Thr A
	Gly	Arg 130	Trp	Thr	Leu	Gly	Ser 135	Ala	Thr	Cys	Lys	Val 140	Val	Arg	Tyr	Phe
20	Gln 145	Tyr	Leu	Thr	Pro	Gly 150	Val	Gln	Ile	Tyr	Val 155	Leu	Leu	Ser	Ile	Cys 160
	Ile	Asp	Arg	Phe	Tyr 165	Thr	Ile	Val	Tyr	Pro 170	Leu	Ser	Phe	Lys	Val 175	Ser
	Arg	Glu	Lys	Ala 180	Lys	Lys	Met	Ile	Ala 185	Ala	Ser	Trp	Ile	Phe 190	Asp	Ala
25	Gly	Phe	Val 195	Thr	Pro	Val	Leu	Phe 200	Phe	Tyr	Gly	Ser	Asn 205	Trp	Asp	Ser
	His	Cys 210	Asn	Tyr	Phe	Leu	Pro 215	Ser	Ser	Trp	Glu	Gly 220	Thr	Ala	Tyr	Thr
30	Val 225	Ile	His	Phe	Leu	Val 230	Gly	Phe	Val	Ile	Pro 235	Ser	Val	Leu	Ile	11e 240
	Leu	Phe	Tyr	Gln	Lys 245	Val	Ile	Lys	Tyr	Ile 250	Trp	Arg	Ile	Gly	Thr 255	Asp
	Gly	Arg	Thr	Val 260	Arg	Arg	Thr	Met	Asn 265	Ile	Val	Pro	Arg	Thr 270	Lys	Val
35	Lys	Thr	Ile 275	Lys	Met	Phe	Leu	Ile 280	Leu	Asn	Leu	Leu	Phe 285	Leu	Leu	Ser

		Trp	Leu 290	Pro	Phe	His	Val	Ala 295	Gln	Leu	Trp	His	Pro 300	His	Glu	Gln	Asp	
		Tyr 305	Lys	Lys	Ser	Ser	Leu 310	Val	Phe	Thr	Ala	Ile 315	Thr	Trp	Ile	Ser	Phe 320	
5		Ser	Ser	Ser	Ala	Ser 325	Lys	Pro	Thr	Leu	Tyr 330	Ser	Ile	Tyr	Asn	Ala 335	Asn	
		Phe	Arg	Arg	Gly 340	Met	Lys	Glu	Thr	Phe 345	Cys	Met	Ser	Ser	Met 350	Lys	Cys	
10		Tyr	Arg	Ser 355	Asn	Ala	Tyr	Thr	Ile 360	Thr	Thr	Ser	Ser	Arg 365	Met	Ala	Lys	
		Lys	Asn 370	Tyr	Val	Gly	Ile	Ser 375	Glu	Ile	Pro	Ser	Met 380	Ala	Lys	Thr	Ile	
		Thr 385	Lys	Asp	Ser	Ile	Tyr 390	Asp	Ser	Phe	Asp	Arg 395	Glu	Ala	Lys	Glu	Lys 400	
15		Lys	Leu	Ala	Trp	Pro 405	Ile	Asn	Ser	Asn	Pro 410	Pro	Asn	Thr	Phe	Val 415		
	(98)) SE	QUEN	FOR CE CI	HARA	CTER	ISTI	CS:	•					·. .~.			
20		٠	(B (C) TY) ST	PE: RAND POLO	nucl EDNE	eic SS:	acid sing										
	•				ULE													
25	·				ENCE					Q ID	NO:	97:						30
25	GGAA				FOR													
	(22)) SE	QUEN	CE C	HARA	CTER	ISTI	CS:	٠								
30			(B) TY	PE: RAND	nucl EDNE	eic SS:	acid sing	l						-			
		(i	i) M	OLEC	ULE	TYPE	: DN	IA (g	enom	ic)				•				
į		(x	:i) S	EQUE	NCE	DESC	RIPI	: NOI	SEÇ	ID	NO:9	8:						
	CTGG																	31
35	(100				N FC													
		(i	.) SE	:QUEN	ICE C	:HARA	CTEF	CISTI	CS:									

PCT/US99/23938

77

- (A) LENGTH: 1842 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- 5 (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

ATGGGGCCCA CCCTAGCGGT TCCCACCCC TATGGCTGTA TTGGCTGTAA GCTACCCCAG 60 CCAGAATACC CACCGGCTCT AATCATCTTT ATGTTCTGCG CGATGGTTAT CACCATCGTT 120 GTAGACCTAA TCGGCAACTC CATGGTCATT TTGGCTGTGA CGAAGAACAA GAAGCTCCGG 180 AATTCTGGCA ACATCTTCGT GGTCAGTCTC TCTGTGGCCG ATATGCTGGT GGCCATCTAC CCATACCCTT TGATGCTGCA TGCCATGTCC ATTGGGGGCT GGGATCTGAG CCAGTTACAG 300 TGCCAGATGG TCGGGTTCAT CACAGGGCTG AGTGTGGTCG GCTCCATCTT CAACATCGTG GCAATCGCTA TCAACCGTTA CTGCTACATC TGCCACAGCC TCCAGTACGA ACGGATCTTC 420 AGTGTGCGCA ATACCTGCAT CTACCTGGTC ATCACCTGGA TCATGACCGT CCTGGCTGTC 480 CTGCCCAACA TGTACATTGG CACCATCGAG TACGATCCTC GCACCTACAC CTGCATCTTC 15 AACTATCTGA ACAACCCTGT CTTCACTGTT ACCATCGTCT GCATCCACTT CGTCCTCCCT 600 CTCCTCATCG TGGGTTTCTG CTACGTGAGG ATCTGGACCA AAGTGCTGGC GGCCCGTGAC · 660 CCTGCAGGGC AGAATCCTGA CAACCAACTT GCTGAGGTTC GCAATTTTCT AACCATGTTT 720 GTGATCTTCC TCCTCTTTGC AGTGTGCTGG TGCCCTATCA ACGTGCTCAC TGTCTTGGTG 780 GCTGTCAGTC CGAAGGAGAT GGCAGGCAAG ATCCCCAACT GGCTTTATCT TGCAGCCTAC 840 TTCATAGCCT ACTTCAACAG CTGCCTCAAC GCTGTGATCT ACGGGCTCCT CAATGAGAAT 900 TTCCGAAGAG AATACTGGAC CATCTTCCAT GCTATGCGGC ACCCTATCAT ATTCTTCCCT 960 GGCCTCATCA GTGATATTCG TGAGATGCAG GAGGCCCGTA CCCTGGCCCG CGCCCGTGCC 1020 CATGCTCGCG ACCAAGCTCG TGAACAAGAC CGTGCCCATG CCTGTCCTGC TGTGGAGGAA ACCCCGATGA ATGTCCGGAA TGTTCCATTA CCTGGTGATG CTGCAGCTGG CCACCCCGAC 1140 CGTGCCTCTG GCCACCCTAA GCCCCATTCC AGATCCTCCT CTGCCTATCG CAAATCTGCC 1200 TCTACCCACC ACAAGTCTGT CTTTAGCCAC TCCAAGGCTG CCTCTGGTCA CCTCAAGCCT 1260 GTCTCTGGCC ACTCCAAGCC TGCCTCTGGT CACCCCAAGT CTGCCACTGT CTACCCTAAG 1320 CCTGCCTCTG TCCATTTCAA GGGTGACTCT GTCCATTTCA AGGGTGACTC TGTCCATTTC 1380

	AAGCCTGACT CTGTTCATTT CAAGCCTGCT TCCAGCAACC CCAAG	SCCCAT CACTGGCCAC 1440
	CATGTCTCTG CTGGCAGCCA CTCCAAGTCT GCCTTCAGTG CTGCC	CACCAG CCACCCTAAA 1500
	CCCATCAAGC CAGCTACCAG CCATGCTGAG CCCACCACTG CTGAC	CTATCC CAAGCCTGCC 1560
	ACTACCAGCC ACCCTAAGCC CGCTGCTGCT GACAACCCTG AGCTC	CTCTGC CTCCCATTGC 1620
5	5 CCCGAGATCC CTGCCATTGC CCACCCTGTG TCTGACGACA GTGAC	CCTCCC TGAGTCGGCC 1680
	TCTAGCCCTG CCGCTGGGCC CACCAAGCCT GCTGCCAGCC AGCTC	GGAGTC TGACACCATC 1740
	GCTGACCTTC CTGACCCTAC TGTAGTCACT ACCAGTACCA ATGAT	TTACCA TGATGTCGTG 1800
	GTTGTTGATG TTGAAGATGA TCCTGATGAA ATGGCTGTGT GA	1842
	(101) INFORMATION FOR SEQ ID NO:100:	. "
10	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 613 amino acids	
	(B) TYPE: amino acid (C) STRANDEDNESS:	
	(D) TOPOLOGY: not relevant	
15	(ii) MOLECULE TYPE: protein	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:	
	Met Gly Pro Thr Leu Ala Val Pro Thr Pro Tyr 1 5 10	Gly Cys Ile Gly Cys 15
20	Lys Leu Pro Gln Pro Glu Tyr Pro Pro Ala Leu 20 25	Ile Ile Phe Met Phe
	Cys Ala Met Val Ile Thr Ile Val Val Asp Leu 35 40	Ile Gly Asn Ser Met 45
	Val Ile Leu Ala Val Thr Lys Asn Lys Lys Leu 50 55	Arg Asn Ser Gly Asn
25	25 Ile Phe Val Val Ser Leu Ser Val Ala Asp Met 65 70 75	Leu Val Ala Ile Tyr 80
	Pro Tyr Pro Leu Met Leu His Ala Met Ser Ile 85 90	Gly Gly Trp Asp Leu 95
30	Ser Gln Leu Gln Cys Gln Met Val Gly Phe Ile	e Thr Gly Leu Ser Val 110
	Val Gly Ser Ile Phe Asn Ile Val Ala Ile Ala 115 120	a Ile Asn Arg Tyr Cys 125
	Tyr Ile Cys His Ser Leu Gln Tyr Glu Arg Ile 130 135	Phe Ser Val Arg Asn 140

areavour and assistant

	Thr 145	Cys	`Ile	Tyr	Leu	Val 150	Ile	Thr	Trp	Ile	Met 155	Thr	Val	Leu	Ala	Va:
	Leu	Pro	Asn	Met	Tyr 165	Ile	Gly	Thr	Ile	Glu 170	Tyr	Asp	Pro	Arg	Thr 175	Туз
5	Thr	Cys	Ile	Phe 180	Asn	Tyr	Leu	Asn	Asn 185	Pro	Val	Phe	Thr	Val 190	Thr	Ile
	Val	Cys	Ile 195	His	Phe	Val	Leu	Pro 200	Leu	Leu	Ile	·Val	Gly 205	Phe	Cys	туг
10	Val	Arg 210	Ile	Trp	Thr	Lys	Val 215	Leu	Ala	Ala	Arg	Asp 220	Pro	Ala	Gly	Glr
	Asn 225	Pro	Asp	Asn	Gln	Leu 230	Ala	Glu	Val	Arg	Asn 235	Phe	Leu	Thr	Met	Phe 240
	Val	Ile	Phe	Leu	Leu 245	Phe	Ala	Val	Cys	Trp 250	Cys	Pro	Ile	Asn	Val 255	Leu
15	Thr	Val	Leu	Val 260	Ala	Val	Ser	Pro	Lys 265	Glu	Met	Ala	Gly	Lys 270	Ile	Pro
	Asn	Trp	Leu 275	Tyr	Leu	Ala	Ala	Tyr 280	Phe	Ile	Ala	Tyr	Phe 285	Asn	Ser	Cys
20	Leu	Asn 290	Ala	Val	Ile	Tyr	Gly 295	Leu	Leu	Asn	Glu	Asn 300	Phe	Arg	Arg	Glu
	Tyr 305	Trp	Thr	Ile	Phe	His 310	Ala	Met	Arg	His	Pro 315	Ile	Ile	Phe	Phe	Pro 320
	Gly	Leu	Ile	Ser	Asp 325	Ile	Arg	Glu	Met	Gln 330	Glu	Ala	Arg	Thr	Leu 335	Ala
25	Arg	Ala	Arg	Ala 340	His	Ala	Arg	Asp	Gln 345	Ala	Arg	Glu	Gln	Asp 350	Arg	Ala
	His	Ala	Cys 355		Ala	Val	Glu		Thr	Pro	Met		Val 365	_	Asn	Val
30	Pro	Leu 370	Pro	Gly	Asp	Ala	Ala 375	Ala	Gly	His	Pro	Asp 380	Arg	Ala	Ser	Gly
	His 385	Pro	Lys	Pro	His	Ser 390	Arg	Ser	Ser	Ser	Ala 395	Tyr	Arg	Lys	Ser	Ala 400
	Ser	Thr	His	His	Lys 405	Ser	Val	Phe	Ser	His 410	Ser	Lys	Ala	Ala	Ser 415	Gly
35	His	Leu	Lys	Pro 420	Val	Ser	Gly	His	Ser 425	Lys	Pro	Ala	Ser	Gly 430	His	Pro
	Lys	Ser	Ala	Thr	Val	Tyr	Pro	Lys	Pro	Ala	Ser	Val	His	Phe	Lys	Gly

				435					440			•,	•	445				
		Asp	Ser 450	Val	His	Phe	Lys	Gly 455	Asp	Ser	Val	His	Phe 460	Lys	Pro	Asp	Ser	
5		Val 465	His	Phe	Lys	Pro	Ala 470	Ser	Ser	Asn	Pro	Lys 475	Pro	Ile	Thr	Gly	His 480	
		His	Val	Ser	Ala	Gly 485	Ser	His	Ser	Lys	Ser 490	Ala	Phe	Ser	Ala	Ala 495	Thr	
		Ser	His		Lys 500	Pro	Ile	Lys	Pro	Ala 505	Thr	Ser	His	Ala	Glu 510	Pro	Thr	
10		Thr	Ala	Asp 515	Tyr	Pro	Lys	Pro	Ala 520	Thr	Thr	Ser	His	Pro 525	Lys	Pro	Ala	
		Ala	Ala 530	Asp	Asn	Pro		Leu -535	Ser	Ala	Ser	His	Cys 540	Pro	Glu	Ile	Pro	
15	N25	Ala 545	Ile	Ala	His	Pro	Val 550	Ser	Asp	Asp	Ser	Asp 555	Leu	Pro	Glu	Ser	Ala 560	
	.192.	Ser	Ser	Pro	Ala	Ala 565	Gly	Pro	Thr	_	Pro 570	Ala	Ala	Ser	Gln	Leu 575	Glu	
••	· 	Ser	Asp	Thr	Ile 580	Ala	Asp	Leu	Pro	Asp 585		Thr	Val	Val	Thr 590	Thr	Ser	
20		Thr	Asn	Asp 595	Tyr	His	Asp	Val	Val 600	Val	Val	Asp	Val	Glu 605	Asp	Asp	Pro	
		Asp	Glu 610	Met	Ala	Val												
	(102) IN	FORM	OITA	N FO	R SE	Q ID	NO:	101:									
25		(i)	(A) (B) (C)) LEI) TY:) STI	NGTH PE: 1 RAND	: 32 nucle EDNE	CTER: base eic a SS:	e pa acid sing	irs									
30									enom.			•						
		(X)	1) 5.	EQUE	NCE I	DESC:	KIPT.	ION:	SEQ	ŤD 1	NO:1	01:						
	TCCA	AGCT"	rc G	CCAT	GGGA	C AT	AACG	GGAG	CT									32
	(103) IN	FORM	ATIO	N FO	R SE	Q ID	NO:	102:									
35		(i	(A (B) LE	NGTH PE::	: 30 nucl	CTER base eic SS:	e pa acid	irs									

PCT/US99/23938

81

í	n') TOPOI	OGV .	lir	
١	· •	, iopoi	JUGI:	111	10 A T

- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

CGTGAATTCC AAGAATTTAC AATCCTTGCT

30

5 (104) INFORMATION FOR SEQ ID NO:103:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1548 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
- 10 (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

ATGGGACATA ACGGGAGCTG GATCTCTCCA AATGCCAGCG AGCCGCACAA CGCGTCCGGC . 60 GCCGAGGCTG CGGGTGTGAA CCGCAGCGCG CTCGGGGAGT TCGGCGAGGC GCAGCTGTAC 120 CGCCAGTTCA CCACCACCGT GCAGGTCGTC ATCTTCATAG GCTCGCTGCT CGGAAACTTC 15 180 ATGGTGTTAT GGTCAACTTG CCGCACAACC GTGTTCAAAT CTGTCACCAA CAGGTTCATT 240 AAAAACCTGG CCTGCTCGGG GATTTGTGCC AGCCTGGTCT GTGTGCCCTT CGACATCATC 300 CTCAGCACCA GTCCTCACTG TTGCTGGTGG ATCTACACCA TGCTCTTCTG CAAGGTCGTC 360 AAATTTTTGC ACAAAGTATT CTGCTCTGTG ACCATCCTCA GCTTCCCTGC TATTGCTTTG 420 GACAGGTACT ACTCAGTCCT CTATCCACTG GAGAGGAAAA TATCTGATGC CAAGTCCCGT 20 480 GAACTGGTGA TGTACATCTG GGCCCATGCA GTGTGGCCA GTGTCCCTGT GTTTGCAGTA 540 ACCAATGTGG CTGACATCTA TGCCACGTCC ACCTGCACGG AAGTCTGGAG CAACTCCTTG 600 GGCCACCTGG TGTACGTTCT GGTGTATAAC ATCACCACGG TCATTGTGCC TGTGGTGGTG 660 GTGTTCCTCT TCTTGATACT GATCCGACGG GCCCTGAGTG CCAGCCAGAA GAAGAAGGTC 720 ATCATAGCAG CGCTCCGGAC CCCACAGAAC ACCATCTCTA TTCCCTATGC CTCCCAGCGG 25 780 GAGGCCGAGC TGCACGCCAC CCTGCTCTCC ATGGTGATGG TCTTCATCTT GTGTAGCGTG 840 CCCTATGCCA CCCTGGTCGT CTACCAGACT GTGCTCAATG TCCCTGACAC TTCCGTCTTC 900 TTGCTGCTCA CTGCTGTTTG GCTGCCCAAA GTCTCCCTGC TGGCAAACCC TGTTCTCTTT 960 CTTACTGTGA ACAAATCTGT CCGCAAGTGC TTGATAGGGA CCCTGGTGCA ACTACACCAC 30 CGGTACAGTC GCCGTAATGT GGTCAGTACA GGGAGTGGCA TGGCTGAGGC CAGCCTGGAA

	CCCAGCATAC GCTCGGGTAG CCAGCTCCTG GAGATGTTCC ACATTGGGCA GCAGCAG	ATC 1140
	TTTAAGCCCA CAGAGGATGA GGAAGAGAGT GAGGCCAAGT ACATTGGCTC AGCTGAC	TTC 1200
,	CAGGCCAAGG AGATATTTAG CACCTGCCTC GAGGGAGAGC AGGGGCCACA GTTTGCG	CCC 1260
	TCTGCCCCAC CCCTGAGCAC AGTGGACTCT GTATCCCAGG TGGCACCGGC AGCCCCT	GTG 1320
5	GAACCTGAAA CATTCCCTGA TAAGTATTCC CTGCAGTTTG GCTTTGGGCC TTTTGAG	TTG 1380
	CCTCCTCAGT GGCTCTCAGA GACCCGAAAC AGCAAGAAGC GGCTGCTTCC CCCCTTG	GGC 1440
	AACACCCCAG AAGAGCTGAT CCAGACAAAG GTGCCCAAGG TAGGCAGGGT GGAGCGG	AAG 1500
	ATGAGCAGAA ACAATAAAGT GAGCATTTTT CCAAAGGTGG ATTCCTAG	1548
	(105) INFORMATION FOR SEQ ID NO:104:	
10	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 515 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: (D) TOPOLOGY: not relevant 	
15	(ii) MOLECULE TYPE: protein	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:	
	Met Gly His Asn Gly Ser Trp Ile Ser Pro Asn Ala Ser Glu F	ro His
		.5
20	Asn Ala Ser Gly Ala Glu Ala Ala Gly Val Asn Arg Ser Ala I 20 25 30	eu Gly
	Glu Phe Gly Glu Ala Gln Leu Tyr Arg Gln Phe Thr Thr W	al Gln
	Val Val Ile Phe Ile Gly Ser Leu Leu Gly Asn Phe Met Val I 50 55 60	eu Trp
25	Ser Thr Cys Arg Thr Thr Val Phe Lys Ser Val Thr Asn Arg F	Phe Ile 80
	Lys Asn Leu Ala Cys Ser Gly Ile Cys Ala Ser Leu Val Cys V	/al Pro 95
30	Phe Asp Ile Ile Leu Ser Thr Ser Pro His Cys Cys Trp Trp 1	Ile Tyr
	Thr Met Leu Phe Cys Lys Val Val Lys Phe Leu His Lys Val 1 115 120 125	Phe Cys
	Ser Val Thr Ile Leu Ser Phe Pro Ala Ile Ala Leu Asp Arg	Tyr Tyr

	Ser 145	Val	Leu	Tyr	Pro	Leu 150	Glu	Arg	Lys	Ile	Ser 155	Asp	Ala	Lys	Ser	Arg 160
	Glu	Leu	Val	Met	Туг 165	Ile	Trp	Ala	His	Ala 170	Val	Val	Ala	Ser	Val 175	Pro
5	Val	Phe	Ala	Val 180	Thr	Asn	Val	Ala	Asp 185	Ile	Tyr	Ala	Thr	Ser 190	Thr	Cys
	Thr	Glu	Val 195	Trp	Ser	Asn	Ser	Leu 200	Gly	His	Leu	Val	Tyr 205	Val	Leu	Val
10	Tyr	Asn 210	Ile	Thr	Thr	Val	Ile 215	Val	Pro	Val	Val	Val 220	Val	Phe	Leu	Phe
	Leu 225	Ile	Leu	Ile	Arg	Arg 230	Ala	Leu	Ser	Ala	Ser 235	Gln	Lys	Lys	Lys	Val 240
	Ile	Ile	Ala	Ala	Leu 245	Arg	Thr	Pro	Gln	Asn 250	Thr	Ile	Ser	Ile	Pro 255	Tyr
15	Ala	Ser	Gln	Arg 260	Glu	Ala	Glu	Leu	His 265	Ala	Thr	Leu	Leu	Ser 270	Met	Val
	Met	Val	Phe 275	Ile	Leu	Cys	Ser	Val 280	Pro	Tyr	Ala	Thr	Leu 285	Val	Val	Tyr
20	Gln	Thr 290	Val	Leu	Asn	Val	Pro 295	Asp	Thr	Ser	Val	Phe 300	Leu	Leu	Leu	Thr
	Ala 305	Val	Trp	Leu	Pro	Lys 310	Val	Ser	Leu	Leu	Ala 315	Asn	Pro	Val	Leu	Phe 320
	Leu	Thr	Val	Asn	Lys 325	Ser	Val	Arg	Lys	Cys 330	Leu	Ile	Gly	Thr	Leu 335	Val
25	Gln	Leu	His	His 340	Arg	Tyr	Ser	Arg	Arg 345	Asn	Val	Val	Ser	Thr 350	Gly	Ser
	Gly	Met	Ala 355	Glu	Ala	Ser	Leu	Glu 360	Pro	Ser	Ile	Arg	Ser 365	Gly	Ser	Gln
30	Leu	Leu 370	Glu	Met	Phe	His	Ile 375	Gly	Gln	Gln	Gln	Ile 380	Phe	Lys	Pro	Thr
	Glu 385	Asp	Glu	Glu	Glu	Ser 390	Glu	Ala	Lys	Tyr	Ile 395	Gly	Ser	Ala	Asp	Phe 400
	Gln	Ala	Lys	Glu	Ile 405	Phe	Ser	Thr	Cys	Leu 410	Glu	Gly	Glu	Gln	Gly 415	Pro
35	Gln	Phe	Ala	Pro 420	Ser	Ala	Pro	Pro	Leu 425	Ser	Thr	Val	Asp	Ser 430	Val	Ser
	Gln	Val	Ala	Pro	Ala	Ala	Pro	Val	Glu	Pro	Glu	Thr	Phe	Pro	Asp	Lys

			435			,	440					443				
	Т	yr Ser 450	Leu (Gln Pł	e Gly	Phe 455	Gly	Pro	Phe	Glu	Leu 460	Pro	Pro	Gln	Trp	
5		eu Ser 65	Glu '	Thr Ai	g Asn 470		Lys	Lys	Arg	Leu 475		Pro	Pro	Leu	Gly 480	
	A	sn Thr	Pro (Glu Gl 48		Ile	Gln	Thr	Lys 490	Val	Pro	Lys	Val	Gly 495	Arg	
	V	al Glu	_	Lys Me 500	et Ser	Arg		Asn 505	Lys	Val	Ser	Ile	Phe 510	Pro	Lys	
10	ν	al Asp	Ser 515										• • •			
•	(106)	INFORM	ATION	FOR S	EQ ID	NO:	105:									
15	•	(B (C	LENG TYP:	E CHAIGTH: 2 E: nuc ANDEDI OLOGY	9 bas cleic NESS:	e pa acid sing	irs									
		(ii) M (xi) S	OLECU	LE T YI	PE: DN	A (g			NO - 1	05.		٠				
20	aarar 1	(XI) S.					SEQ	10		03.					•	29
20		INFORM					106:									2
25		(B (C) LEN) TYP) STR	E CHAI GTH: : E: nuc ANDEDI OLOGY	30 bas cleic NESS:	e pa acid sing	irs							· .		
		(ii) M	oLECU	LE TY	PE: DN	IA (g	enom	ic)								
		(xi) S	EQUEN	CE DE	SCRIPI	ION:	SEQ	ID:	NO:1	06:						
	GGAGG!	ATCCA G	GAAAC	CTTA	GCCGA	GTCC	!									30
30	(108)	INFORM	ATION	FOR	SEQ II	NO:	107:									
		(B) LEN	E CHA GTH: E: nu ANDED	1164 k cleic	ase acid	pair l	s		•						
35				OLOGY		_	,									
		(ii) M	OLECU	LE TY	PE: DN	JA (g	enom	ic)								

85

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

	ATGAATCGGC	ACCATCTGCA	GGATCACTTT	CTGGAAATAG	ACAAGAAGAA	CTGCTGTGTG	60
	TTCCGAGATG	ACTTCATTGC	CAAGGTGTTG	CCGCCGGTGT	TGGGGCTGGA	GTTTATCTTT	120
	GGGCTTCTGG	GCAATGGCCT	TGCCCTGTGG	ATTTTCTGTT	TCCACCTCAA	GTCCTGGAAA	180
5	TCCAGCCGGA	TTTTCCTGTT	CAACCTGGCA	GTAGCTGACT	TTCTACTGAT	CATCTGCCTG	240
	CCGTTCGTGA	TGGACTACTA	TGTGCGGCGT	TCAGACTGGA	ACTTTGGGGA	CATCCCTTGC	300
	CGGCTGGTGC	TCTTCATGTT	TGCCATGAAC	CGCCAGGGCA	GCATCATCTT	CCTCACGGTG	360
	GTGGCGGTAG	ACAGGTATTT	CCGGGTGGTC	CATCCCCACC	ACGCCCTGAA	CAAGATCTCC	420
	AATTGGACAG	CAGCCATCAT	CTCTTGCCTT	CTGTGGGGCA	TCACTGTTGG	CCTAACAGTC	480
10	CACCTCCTGA	AGAAGAAGTT	GCTGATCCAG	AATGGCCCTG	CAAATGTGTG	CATCAGCTTC	540
	AGCATCTGCC	ATACCTTCCG	GTGGCACGAA	GCTATGTTCC	TCCTGGAGTT	CCTCCTGCCC	600
	CTGGGCATCA	TCCTGTTCTG	CTCAGCCAGA	ATTATCTGGA	GCCTGCGGCA	GAGACAAATG	660
	GACCGGCATG	CCAAGATCAA	GAGAGCCATC	ACCTTCATCA	TGGTGGTGGC	CATCGTCTTT	720
	GTCATCTGCT	TCCTTCCCAG	CGTGGTTGTG	CGGATCCGCA	TCTTCTGGCT	CCTGCACACT	780
15	TCGGGCACGC	AGAATTGTGA	AGTGTACCGC	TCGGTGGACC	TGGCGTTCTT	TATCACTCTC	840
	AGCTTCACCT	ACATGAACAĠ	CATGCTGGAC	CCCGTGGTGT	ACTACTTCTC	CAGCCCATCC	900
	TTTCCCAACT	TCTTCTCCAC	TTTGATCAAC	CGCTGCCTCC	AGAGGAAGAT	GACAGGTGAG	960
	CCAGATAATA	ACCGCAGCAC	GAGCGTCGAG	CTCACAGGGG	ACCCCAACAA	AACCAGAGGC	1020
	GCTCCAGAGG	CGTTAATGGC	CAACTCCGGT	GAGCCATGGA	GCCCCTCTTA	TCTGGGCCCA	1080
20	ACCTCAAATA	ACCATTCCAA	GAAGGGACAT	TGTCACCAAG	AACCAGCATC	TCTGGAGAAA	1140
	CAGTTGGGCT	GTTGCATCGA	GTAA	•			1164

(109) INFORMATION FOR SEQ ID NO:108:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 387 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:

- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:
- 30 Met Asn Arg His His Leu Gln Asp His Phe Leu Glu Ile Asp Lys

	1				5			: · ·		10	-4	٠			15	
	Asn	Cys	Cys	Val 20	Phe	Arg	Asp	Asp	Phe 25	Ile	Ala	Lys	Val	Leu 30	Pro	Pro
5	Val	Leu	Gly 35	Leu	Glu	Phe	Ile	Phe 40	Gly	Leu	Leu	Gly	Asn 45	Gly	Leu	Ala
	Leu	Trp 50	Ile	Phe	Cys	Phe	His 55	Leu	Lys	Ser	Trp	Lys 60	Ser	Ser	Arg	Ile
		Leu	Phe	Asn	Leu	Ala 70	Val	Ala	Asp	Phe	Leu 75	Leu	Ile	Ile	Cys	Leu 80
10	Pro	Phe	Val	Met	Asp 85	Tyr	Tyr	Val	Arg	Arg 90	Ser	Asp	Trp	Asn	Phe 95	Gly
	Asp	Ile	Pro	Cys 100	Arg	Leu	Val		Phe 105	Met	Phe	Ala	Met	Asn 110		Gln
15	Gly	Ser	Ile 115	Ile	Phe	Leu	Thr	Val 120	Val	Ala	Val	Asp	Arg 125	Tyr	Phe	Arg
	Val	Val 130	His	Pro	His	His	Ala 135	Leu	Asn	Lys	Ile	Ser 140	Asn	Trp	Thr	Ala
	Ala 145	Ile	Ile	Ser	Cys	Leu 150	Leu	Trp	Gly	Ile	Thr 155	Val	Gly	Leu	Thr	Val 160
20	His	Leu	Leu	Lys	Lys 165	Lys	Leu	Leu	Ile	Gln 170	Asn	Gly	Pro	Ala	Asn 175	Val
	Cys	Ile	Ser	Phe 180	Ser	Ile	Cys	His	Thr 185	Phe	Arg	Trp	His	Glu 190	Ala	Met
25	Phe	Leu	Leu 195	Glu	Phe	Leu	Leu	Pro 200	Leu	Gly	Ile	Ile	Leu 205	Phe	Cys	Ser
	Ala	Arg 210	Ile	Ile	Trp	Ser	Leu 215	Arg	Gln	Arg	Gln	Met 220	Asp	Arg	His	Ala
	Lys 225	Ile	Lys	Arg	Ala	Ile 230	Thr	Phe	Ile	Met	Val 235	Val	Ala	Ile	Val	Phe 240
30	Val	Ile	Cys	Phe	Leu 245	Pro	Ser	Val	Val	Val 250		Ile	Arg	Ile	Phe 255	Trp
	Leu	Leu	His	Thr 260		Gly	Thr	Gln	Asn 265		Glu	Val	Tyr	Arg 270	Ser	Val
35	Asp	Leu	Ala 275		Phe	Ile	Thr	Leu 280		Phe	Thr	Tyr	Met 285		Ser	Met
	Leu	Asp 290	Pro	Val	Val	Tyr	Tyr 295		Ser	Ser	Pro	Ser 300		Pro	Asn	Phe

מאופרעיים אות בייים מייים וייים בייים וייים בייים וייים בייים בייי

87

	Phe Ser Thr Leu Ile Asn Arg Cys Leu Gln Arg Lys Met Thr Gly Glu 305 310 315 320
	Pro Asp Asn Asm Arg Ser Thr Ser Val Glu Leu Thr Gly Asp Pro Asn 325 330 335
5	Lys Thr Arg Gly Ala Pro Glu Ala Leu Met Ala Asn Ser Gly Glu Pro 340 345 350
	Trp Ser Pro Ser Tyr Leu Gly Pro Thr Ser Asn Asn His Ser Lys Lys 355 360 365
10	Gly His Cys His Gln Glu Pro Ala Ser Leu Glu Lys Gln Leu Gly Cys 370 375 380
•	Cys Ile Glu 385
	(110) INFORMATION FOR SEQ ID NO:109:
15	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 37 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear
	(ii) MOLECULE TYPE: DNA (genomic)
20	(iv) ANTI-SENSE: NO
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:
	ACCATGGCTT GCAATGGCAG TGCGGCCAGG GGGCACT
	(111) INFORMATION FOR SEQ ID NO:110:
25	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear
	(ii) MOLECULE TYPE: DNA (genomic)
30	(iv) ANTI-SENSE: YES
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:
	CGACCAGGAC AAACAGCATC TTGGTCACTT GTCTCCGGC
	(112) INFORMATION FOR SEQ ID NO:111:
35	(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 39 base pairs(B) TYPE: nucleic acid

A CONTRACTOR

	(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
5	(iv) ANTI-SENSE: NO	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:	
	GACCAAGATG CTGTTTGTCC TGGTCGTGGT GTTTGGCAT	39
	(113) INFORMATION FOR SEQ ID NO:112:	
10	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 35 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
15	(iv) ANTI-SENSE: YES	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:	
	CGGAATTCAG GATGGATCGG TCTCTTGCTG CGCCT	35
	(114) INFORMATION FOR SEQ ID NO:113:	
20	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1212 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
25	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:	
	ATGGCTTGCA ATGGCAGTGC GGCCAGGGGG CACTTTGACC CTGAGGACTT GAACCTGACT	60
	GACGAGGCAC TGAGACTCAA GTACCTGGGG CCCCAGCAGA CAGAGCTGTT CATGCCCATC	120
	TGTGCCACAT ACCTGCTGAT CTTCGTGGTG GGCGCTGTGG GCAATGGGCT GACCTGTCTG	180
	GTCATCCTGC GCCACAAGGC CATGCGCACG CCTACCAACT ACTACCTCTT CAGCCTGGCC	240
30	GTGTCGGACC TGCTGGTGCT GCTGGTGGGC CTGCCCCTGG AGCTCTATGA GATGTGGCAC	300
	AACTACCCCT TCCTGCTGGG CGTTGGTGGC TGCTATTTCC GCACGCTACT GTTTGAGATG	360
	GTCTGCCTGG CCTCAGTGCT CAACGTCACT GCCCTGAGCG TGGAACGCTA TGTGGCCGTG	420
	GTGCACCCAC TCCAGGCCAG GTCCATGGTG ACGCGGGCCC ATGTGCGCCG AGTGCTTGGG	480

	GCCGTCTGGG GTCTTGCCAT GCTCTGCTCC CTGCCCAACA CCAGCCTGCA CGGCATCCGG 54	0
	CAGCTGCACG TGCCCTGCCG GGGCCCAGTG CCAGACTCAG CTGTTTGCAT GCTGGTCCGC 60	0
	CCACGGGCCC TCTACAACAT GGTAGTGCAG ACCACCGCGC TGCTCTTCTT CTGCCTGCCC 66	0
	ATGGCCATCA TGAGCGTGCT CTACCTGCTC ATTGGGCTGC GACTGCGGCG GGAGAGGCTG 72	0
5	CTGCTCATGC AGGAGGCCAA GGGCAGGGGC TCTGCAGCAG CCAGGTCCAG ATACACCTGC 78	0
	AGGCTCCAGC AGCACGATCG GGGCCGGAGA CAAGTGACCA AGATGCTGTT TGTCCTGGTC 84	0
	GTGGTGTTTG GCATCTGCTG GGCCCCGTTC CACGCCGACC GCGTCATGTG GAGCGTCGTG 90	0
	TCACAGTGGA CAGATGGCCT GCACCTGGCC TTCCAGCACG TGCACGTCAT CTCCGGCATC 96	0
	TTCTTCTACC TGGGCTCGGC GGCCAACCCC GTGCTCTATA GCCTCATGTC CAGCCGCTTC 102	0
10	CGAGAGACCT TCCAGGAGGC CCTGTGCCTC GGGGCCTGCT GCCATCGCCT CAGACCCCGC 108	0
	CACAGCTCCC ACAGCCTCAG CAGGATGACC ACAGGCAGCA CCCTGTGTGA TGTGGGCTCC 114	0
	CTGGGCAGCT GGGTCCACCC CCTGGCTGGG AACGATGGCC CAGAGGCGCA GCAAGAGACC 120	0
	GATCCATCCT GA	2
	(115) INFORMATION FOR SEQ ID NO:114:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 403 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: (D) TOPOLOGY: not relevant 	
20	(ii) MOLECULE TYPE: protein	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:	
	Met Ala Cys Asn Gly Ser Ala Ala Arg Gly His Phe Asp Pro Glu Asp 1 5 10 15	
25	Leu Asn Leu Thr Asp Glu Ala Leu Arg Leu Lys Tyr Leu Gly Pro Gln 20 25 30	
	Gln Thr Glu Leu Phe Met Pro Ile Cys Ala Thr Tyr Leu Leu Ile Phe 35 40 45	
	Val Val Gly Ala Val Gly Asn Gly Leu Thr Cys Leu Val Ile Leu Arg 50 55 60	
30	His Lys Ala Met Arg Thr Pro Thr Asn Tyr Tyr Leu Phe Ser Leu Ala 65 70 75 80	
	Val Ser Asp Leu Leu Val Leu Val Gly Leu Pro Leu Glu Leu Tyr	

	GIU	Met	irp,	100	ASII	ıyı	PIO	Pne,	105	Leu	GIY	vaı	GIÀ	110	Cys	ıyr
	Phe	Arg	Thr 115	Leu	Leu	Phe	Glu	Met 120	Val	Cys	Leu	Ala	Ser 125	Val	Leu	Asn
5	Val	Thr 130	Ala	Leu	Ser	Val	Glu 135	Arg	Tyr	Val	Ala	Val 140	Val	His	Pro	Leu
	Gln 145	Ala	Arg	Ser	Met	Val 150	Thr	Arg	Ala		Val 155		Arg	Val	Leu	Gly 160
10	Ala	Val	Trp	Gly	Leu 165	Ala	Met	Leu	Cys	Ser 170	Leu	Pro	Asn	Thr	Ser 175	Leu
	His	Gly	Ile	Arg 180	Gln	Leu	His	Val	Pro 185	Cys	Arg	Gly	Pro	Val 190		Asp
	Ser	Ala	Val 195	Суз	Met	Leu	Val	Arg 200	Pro	Arg	Ala	Leu	Tyr 205	Asn	Met	Val
15	Val	Gln 210	Thr	Thr	Ala	Leu	Leu 215	Phe	Phe	Cys	Leu	Pro 220	Met	Ala	Ile	Met
	225					230			:		235	,	Arg			240
20					245					250			Ala		255	
				260					265				Arg	270		
25			275					280				_	Ile 285		_	
25		290					295					300	Ser			
	305					310					315		Ile		-	320
30					325					330			Tyr		335	
				340					345				Cys	350		
35	-	-	355					360					365 Leu			
		370					375					380				Thr

PCT/US99/23938

	·	
	390 395	100
	Asp Pro Ser	
	(116) INFORMATION FOR SEQ ID NO:115:	
5	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
10	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:	
	GGAAGCTTCA GGCCCAAAGA TGGGGAACAT	30
	(117) INFORMATION FOR SEQ ID NO:116:	30
15	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	e de la companya de l
٠	(ii) MOLECULE TYPE: DNA (genomic)	' ".
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:	
	GTGGATCCAC CCGCGGAGGA CCCAGGCTAG	30
	(118) INFORMATION FOR SEQ ID NO:117:	
25	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1098 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	· .
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:	
30	ATGGGGAACA TCACTGCAGA CAACTCCTCG ATGAGCTGTA CCATCGACCA TACCATCCAC	60
	CAGACGCTGG CCCCGGTGGT CTATGTTACC GTGCTGGTGG TGGGCTTCCC GGCCAACTGC	120
	CTGTCCCTCT ACTTCGGCTA CCTGCAGATC AAGGCCCGGA ACGAGCTGGG CGTGTACCTG	180
	TGCAACCTGA CGGTGGCCGA CCTCTTCTAC ATCTGCTCGC TGCCCTTCTG GCTGCAGTAC	240
	GTGCTGCAGC ACGACAACTG GTCTCACGGC GACCTGTCCT GCCAGGTGTG CGGCATCCTC	300
35	CTGTACGAGA ACATCTACAT CAGCGTGGGC TTCCTCTGCT GCATCTCCGT GGACCGCTAC	360

	CTGGCTGTGG	CCCATCCCTT	CCGCTTCCAC	CAGTTCCGGA	CCCTGAAGGC	GGCCGTCGGC	420
	GTCAGCGTGG	TCATCTGGGC	CAAGGAGCTG	CTGACCAGCA	TCTACTTCCT	GATGCACGAG	480
	GAGGTCATCG	AGGACGAGAA	CCAGCACCGC	GTGTGCTTTG	AGCACTACCC	CATCCAGGCA	540
	TGGCAGCGCG	CCATCAACTA	CTACCGCTTC	CTGGTGGGCT	TCCTCTTCCC	CATCTGCCTG	600
5	CTGCTGGCGT	CCTACCAGGG	CATCCTGCGC	GCCGTGCGCC	GGAGCCACGG	CACCCAGAAG	660
	AGCCGCAAGG	ACCAGATCCA	GCGGCTGGTG	CTCAGCACCG	TGGTCATCTT	CCTGGCCTGC	720
	TTCCTGCCCT	ACCACGTGTT	GCTGCTGGTG	CGCAGCGTCT	GGGAGGCCAG	CTGCGACTTC	780
	GCCAAGGGCG	TTTTCAACGC	CTACCACTTC	TCCCTCCTGC	TCACCAGCTT	CAACTGCGTC	840
	GCCGACCCCG	TGCTCTACTG	CTTCGTCAGC	GAGACCACCC	ACCGGGACCT	GGCCCGCCTC	900
10	CGCGGGGCCT	GCCTGGCCTT	CCTCACCTGC	TCCAGGACCG	GCCGGGCCAG	GGAGGCCTAC	960
	CCGCTGGGTG	CCCCGAGGC	CTCCGGGAAA	AGCGGGGCCC	AGGGTGAGGA	GCCCGAGCTG	1020
	TTGACCAAGC	TCCACCCGGC	CTTCCAGACC	CCTAACTCGC	CAGGGTCGGG	CGGGTTCCCC	1080
	ACGGGCAGGT	TGGCCTAG				•	1098

(119) INFORMATION FOR SEQ ID NO:118:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 365 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- 20 (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

Met Gly Asn Ile Thr Ala Asp Asn Ser Ser Met Ser Cys Thr Ile Asp 1 5 10 15

His Thr Ile His Gln Thr Leu Ala Pro Val Val Tyr Val Thr Val Leu 25 25 30

Val Val Gly Phe Pro Ala Asn Cys Leu Ser Leu Tyr Phe Gly Tyr Leu 35 40 45

Gln Ile Lys Ala Arg Asn Glu Leu Gly Val Tyr Leu Cys Asn Leu Thr 50 55 60

Val Ala Asp Leu Phe Tyr Ile Cys Ser Leu Pro Phe Trp Leu Gln Tyr 65 70 75 80

Val Leu Gln His Asp Asn Trp Ser His Gly Asp Leu Ser Cys Gln Val 85 90 95

	1	Cys	Gly	Ile	Leu 100	Leu	Tyr	Glu	Asn	Ile 105	Tyr	Ile	Ser	Val	Gly 110	Phe	Leu
		Cys	Cys	Ile 115	Ser	Val	Asp	Arg	Tyr 120	Leu	Ala	Val	Ala	His 125	Pro	Phe	Arg
5		Phe	His 130	Gln	Phe	Arg	Thr	Leu 135	Lys	Ala	Ala	Val	Gly 140	Val	Ser	Val	Val
		Ile 145	Trp	Ala	Lys	Glu	Leu 150	Leu	Thr	Ser	Ile	Tyr 155	Phe	Leu	Met	His	Glu 160
10		Glu	Val	Ile	Glu	Asp 165	Glu	Asn	Gln	His	Arg 170	Val	Cys	Phe	Glu	His 175	Tyr
		Pro	Ile	Gln	Ala 180	Trp	Gln	Arg	Ala	Ile 185	Asn	Tyr	Tyr	Arg	Phe 190	Leu	Val
		Gly	Phe	Leu 195	Phe	Pro	Ile	Cys	Leu 200	Leu	Leu	Ala	Ser	Tyr 205	Gln	Gly	Ile
15		Leu	Arg 210	Ala	Val	Arg	Arg	Ser 215	His	Gly	Thr	Gln	Lys 220	Ser	Arg	Lys	Asp
		Gln 225	Ile	Gln	Arg	Leu	Val 230	Leu	Ser	Thr	Val	Val 235	Ile	Phe	Leu	Ala	Cys 240
20		Phe	Leu	Pro	Tyr	His 245	Val	Leu	Leu	Leu	Val 250	Arg	Ser	Val	Trp	Glu 255	Ala
		Ser	Cys	Asp	Phe 260	Ala	Lys	Gly	Val	Phe 265	Asn	Ala	Tyr	His	Phe 270	Ser	Leu
		Leu	Leu	Thr 275	Ser	Phe	Asn	Cys	Val 280	Ala	Asp	Pro	Val	Leu 285	Tyr	Cys	Phe
25		Val	Ser 290	Glu	Thr	Thr	His	Arg 295	Asp	Leu	Ala	Arg	Leù 300	Arg	Gly	Ala	Cys
		Leu 305	Ala	Phe	Leu	Thr	Cys 310	Ser	Arg	Thr	Gly	Arg 315	Ala	Arg	Glu	Ala	Tyr 320
30		Pro	Leu	Gly	Ala	Pro 325	Glu	Ala	Ser	Gly	Lys 330	Ser	Gly	Ala	Gln	Gly 335	Glu
		Glu	Pro	Glu	Leu 340	Leu	Thr	Lys	Leu	His 345	Pro	Ala	Phe	Gln	Thr 350	Pro	Asn
		Ser	Pro	Gly 355	Ser	Gly	Gly	Phe	Pro 360	Thr	Gly	Arg	Leu	Ala 365			
35	(120)	TNE	COPM2	TON	T EOE	9 SEC	מד נ	NO · 1	19.								

- 35 (120) INFORMATION FOR SEQ ID NO:119:
 - (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 26 base pairs

	(B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
5	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:	
	GACCTCGAGT CCTTCTACAC CTCATC	26
	(121) INFORMATION FOR SEQ ID NO:120:	
10	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:	
15	TGCTCTAGAT TCCAGATAGG TGAAAACTTG	30
	(122) INFORMATION FOR SEQ ID NO:121:	
20	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1416 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:	
	ATGGATATTC TTTGTGAAGA AAATACTTCT TTGAGCTCAA CTACGAACTC CCTAATGCAA	60
25	TTAAATGATG ACAACAGGCT CTACAGTAAT GACTTTAACT CCGGAGAAGC TAACACTTCT	120
	GATGCATTTA ACTGGACAGT CGACTCTGAA AATCGAACCA ACCTTTCCTG TGAAGGGTGC	180
	CTCTCACCGT CGTGTCTCTC CTTACTTCAT CTCCAGGAAA AAAACTGGTC TGCTTTACTG	240
	ACAGCCGTAG TGATTATTCT AACTATTGCT GGAAACATAC TCGTCATCAT GGCAGTGTCC	300
	CTAGAGAAAA AGCTGCAGAA TGCCACCAAC TATTTCCTGA TGTCACTTGC CATAGCTGAT	360
30	ATGCTGCTGG GTTTCCTTGT CATGCCCGTG TCCATGTTAA CCATCCTGTA TGGGTACCGG	420
	TGGCCTCTGC CGAGCAAGCT TTGTGCAGTC TGGATTTACC TGGACGTGCT CTTCTCCACG	480
	GCCTCCATCA TGCACCTCTG CGCCATCTCG CTGGACCGCT ACGTCGCCAT CCAGAATCCC	540
	ATCCACCACA GCCGCTTCAA CTCCAGAACT AAGGCATTTC TGAAAATCAT TGCTGTTTGG	600

BYIGUALID -IMA WASHINGT

				95			
	ACCATATCAG T	AGGTATATC	CATGCCAATA	CCAGTCTTTG	GGCTACAGGA	CGATTCGAAG	660
	GTCTTTAAGG A	GGGGAGTTG	CTTACTCGCC	GATGATAACT	TTGTCCTGAT	CGGCTCTTTT	720
	GTGTCATTTT T	CATTCCCTT	AACCATCATG	GTGATCACCT	ACTTTCTAAC	TATCAAGTCA	780
	CTCCAGAAAG A	AGCTACTTT	GTGTGTAAGT	GATCTTGGCA	CACGGGCCAA	ATTAGCTTCT	840
5	TTCAGCTTCC T	CCCTCAGAG	TTCTTTGTCT	TCAGAAAAGC	TCTTCCAGCG	GTCGATCCAT	900
	AGGGAGCCAG G	GTCCTACAC	AGGCAGGAGG	ACTATGCAGT	CCATCAGCAA	TGAGCAAAAG	960
	GCATGCAAGG T	GCTGGGCAT	CGTCTTCTTC	CTGTTTGTGG	TGATGTGGTG	CCCTTTCTTC	1020
	ATCACAAACA T	CATGGCCGT	CATCTGCAAA	GAGTCCTGCA	ATGAGGATGT	CATTGGGGCC	1080
	CTGCTCAATG T	GTTTGTTTG	GATCGGTTAT	CTCTCTTCAG	CAGTCAACCC	ACTAGTCTAC	1140
. 10	ACACTGTTCA A	CAAGACCTA	TAGGTCAGCC	TTTTCACGGT	ATATTCAGTG	TCAGTACAAG	1200
	GAAAACAAAA A	ACCATTGCA	GTTAATTTTA	GTGAACACAA	TACCGGCTTT	GGCCTACAAG	1260
	TCTAGCCAAC T	TCAAATGGG	ACAAAAAAAG	AATTCAAAGC	AAGATGCCAA	GACAACAGAT	1320
	AATGACTGCT C	AATGGTTGC	TCTAGGAAAG	CAGTATTCTG	AAGAGGCTTC	TAAAGACAAT	1380
	AGCGACGGAG TO	GAATGAAAA	GGTGAGCTGT	GTGTGA			1416
15	(123) INFORM	ATION FOR	SEQ ID NO:1	122:			-
	(A (B						
20	• -		: not relev	vant			
	(ii) MOL	ECULE. TYPE	: protein				
	(xi) SEQUENCE	E DESCRIPT	'ION: SEQ II	NO:122:			

Met Asp Ile Leu Cys Glu Glu Asn Thr Ser Leu Ser Ser Thr Thr Asn 1 5 10 15

25 Ser Leu Met Gln Leu Asn Asp Asp Asn Arg Leu Tyr Ser Asn Asp Phe 20 25 30

Asn Ser Gly Glu Ala Asn Thr Ser Asp Ala Phe Asn Trp Thr Val Asp 35 40 45

Ser Glu Asn Arg Thr Asn Leu Ser Cys Glu Gly Cys Leu Ser Pro Ser 30 50 55 60

Cys Leu Ser Leu Leu His Leu Gln Glu Lys Asn Trp Ser Ala Leu Leu 65 70 75 80

	Thr	Ala	Val	Val	Ile 85	Ile	Leu	Thr	Ile	Ala 90	Gly	Asn	Ile	Leu	Val 95	Ile
	Met	Ala	Val	Ser 100	Leu	Glu	Lys	Lys	Leu 105	Gln	Asn	Ala	Thr	Asn 110	Tyr	Phe
5	Leu	Met	Ser 115	Leu	Ala	Ile	Ala	Asp 120	Met	Leu	Leu	Gly	Phe 125	Leu	Val	Met
	Pro	Val 130	Ser	Met	Leu	Thr	Ile 135	Leu	Tyr	Gly	Tyr	Arg 140	Trp	Pro	Leu	Pro
10	Ser 145	Lys	Leu	Суз	Ala	Val 150	Trp	Ile	Tyr	Leu	Asp 155	Val	Leu	Phe	Ser	Thr 160
	Ala	Ser	Ile	Met	His 165	Leu	Cys	Ala	Ile	Ser 170	Leu	Asp	Arg	Tyr	Val 175	Ala
	Ile	Gln	Asn	Pro 180	Ile	His	His	Ser	Arg 185	Phe	Asn	Ser	Arg	Thr 190	Lys	Ala
15	Phe	Leu	Lys 195	Ile	Ile	Ala	Val	Trp 200	Thr	Ile	Ser	Val	Gly 205	Ile	Ser	Met
	Pro	Ile 210	Pro	Val	Phe	Gly	Leu 215	Gln	Asp	Asp	Ser	Lys 220	Val	Phe	Lys	Glu
20	Gly 225	Ser	Cys	Leu	Leu	Ala 230	Asp	Asp	Asn	Phe	Val 235	Leu	Ile	Gly	Ser	Phe 240
	Val	Ser	Phe	Phe	Ile 245	Pro	Leu	Thr	Ile	Met 250	Val	Ile	Thr	Tyr	Phe 255	Leu
	Thr	Ile	Lys	Ser 260	Leu	Gln	Lys	Glu	Ala 265	Thr	Leu	Cys	Val	Ser 270	-	Leu
25	Gly	Thr	Arg 275	Ala	Lys	Leu	Ala	Şer 280	Phe	Ser	Phe	Leu	Pro 285	Gln	Ser	Ser
	Leu	Ser 290	Ser	Glu	Lys	Leu	Phe 295	Gln	Arg	Ser	Ile	His 300	Arg	Glu	Pro	Gly
30	Ser 305	Tyr	Thr	Gly	Arg	Arg 310	Thr	Met	Gln	Ser	Ile 315	Ser	Asn	Glu	Gln	Lys 320
	Ala	Cys	Lys	Val	Leu 325	Gly	Ile	Val	Phe			Phe		Val	Met 335	Trp
	Cys	Pro	Phe	Phe 340	Ile	Thr	Asn [.]	Ile	Met 345	Ala	Val	Ile	Cys	Lys 350	Glu	Ser
35	Cys	Asn	Glu 355	Asp	Val	Ile	Gly	Ala 360	Leu	Leu	Asn	Val	Phe 365	Val	Trp	Ile
	Gly	Tyr	Leu	Ser	Ser	Ala	Val	Asn	Pro	Leu	Val	Tyr	Thr	Leu	Phe	Asn

PCT/US99/23938

	370	
	Lys Thr Tyr Arg Ser Ala Phe Ser Arg Tyr Ile Gln Cys Gln Tyr 385 390 395	Lys 400
5	Glu Asn Lys Lys Pro Leu Gln Leu Ile Leu Val Asn Thr Ile Pro 405 410 415	Ala
	Leu Ala Tyr Lys Ser Ser Gln Leu Gln Met Gly Gln Lys Lys Asn 420 425 430	Ser
	Lys Gln Asp Ala Lys Thr Thr Asp Asn Asp Cys Ser Met Val Ala 435 440 445	Leu
10	Gly Lys Gln Tyr Ser Glu Glu Ala Ser Lys Asp Asn Ser Asp Gly 450 455 460	Val
	Asn Glu Lys Val Ser Cys Val 465 470	
	(124) INFORMATION FOR SEQ ID NO:123:	٠.
15	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 27 base pairs(B) TYPE: nucleic acid	٠,
	(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
20	(ii) MOLECULE TYPE: DNA (genomic)	· .
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:	,
	GACCTCGAGG TTGCTTAAGA CTGAAGC	27
	(125) INFORMATION FOR SEQ ID NO:124:	
25	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 27 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
30	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:124:	
	ATTTCTAGAC ATATGTAGCT TGTACCG	27
	(126) INFORMATION FOR SEQ ID NO:125:	
35	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1377 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

	ATGGTGAACC	TGAGGAATGC	GGTGCATTCA	TTCCTTGTGC	ACCTAATTGG	CCTATTGGTT	60
	TGGCAATGTG	ATATTTCTGT	GAGCCCAGTA	GCAGCTATAG	TAACTGACAT	TTTCAATACC	120
5	TCCGATGGTG	GACGCTTCAA	ATTCCCAGAC	GGGGTACAAA	ACTGGCCAGC	ACTTTCAATC	180
	GTCATCATAA	TAATCATGAC	AATAGGTGGC	AACATCCTTG	TGATCATGGC	AGTAAGCATG	240
	GAAAAGAAAC	TGCACAATGC	CACCAATTAC	TTCTTAATGT	CCCTAGCCAT	TGCTGATATG	300
	CTAGTGGGAC	TACTTGTCAT	GCCCCTGTCT	CTCCTGGCAA	TCCTTTATGA	TTATGTCTGG	360
	CCACTACCTA	GATATTTGTG	CCCCGTCTGG	ATTTCTTTAG	ATGTTTTATT	TTCAACAGCG	420
10	TCCATCATGC	ACCTCTGCGC	TATATCGCTG	GATCGGTATG	TAGCAATACG	TAATCCTATT	480
	GAGCATAGCC	GTTTCAATTC	GCGGACTAAG	GCCATCATGA	AGATTGCTAT	TGTTTGGGCA	540
	ATTTCTATAG	GTGTATCAGT	TCCTATCCCT	GTGATTGGAC	TGAGGGACGA	AGAAAAGGTG	600
	TTCGTGAACA	ACACGACGTG	CGTGCTCAAC	GACCCAAATT	TCGTTCTTAT	TGGGTCCTTC	660
•	GTAGCTTTCT	TCATACCGCT	GACGATTATG	GTGATTACGT	ATTGCCTGAC	CATCTACGTT	720
15	CTGCGCCGAC	AAGCTTTGAT	GTTACTGCAC	GGCCACACCG	AGGAACCGCC	TGGACTAAGT	780
	CTGGATTTCC	TGAAGTGCTG	CAAGAGGAAT	ACGGCCGAGG	AAGAGAACTC	TGCAAACCCT	840
	AACCAAGACC	AGAACGCACG	CCGAAGAAAG	AAGAAGGAGA	GACGTCCTAG	GGGCACCATG	900
	CAGGCTATCA	ACAATGAAAG	AAAAGCTTCG	AAAGTCCTTG	GGATTGTTTT	CTTTGTGTTT	960
	CTGATCATGT	GGTGCCCATT	TTTCATTACC	AATATTCTGT	CTGTTCTTTG	TGAGAAGTCC	1020
20	TGTAACCAAA	AGCTCATGGA	AAAGCTTCTG	AATGTGTTTG	TTTGGATTGG	CTATGTTTGT	1080
	TCAGGAATCA	ATCCTCTGGT	GTATACTCTG	TTCAACAAAA	TTTACCGAAG	GGCATTCTCC	1140
	AACTATTTGC	GTTGCAATTA	TAAGGTAGAG	AAAAAGCCTC	CTGTCAGGCA	GATTCCAAGA	1200
	GTTGCCGCCA	CTGCTTTGTC	TGGGAGGGAG	CTTAATGTTA	ACATTTATCG	GCATACCAAT	1260
	GAACCGGTGA	TCGAGAAAGC	CAGTGACAAT	GAGCCCGGTA	TAGAGATGCA	AGTTGAGAAT	1320
25	TTAGAGTTAC	CAGTAAATCC	CTCCAGTGTG	GTTAGCGAAA	GGATTAGCAG	TGTGTGA	1377

(127) INFORMATION FOR SEQ ID NO:126:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 458 amino acids
 - (B) TYPE: amino acid

- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

			~					_								
5	Met 1	Val	Asn	Leu	Arg 5	Asn	Ala	Val	His	Ser 10	Phe	Leu	Val	His	Leu 15	
	Gly	Leu	Leu	Val 20	Trp	Gln	Cys	Asp	Ile 25	Ser	Val	Ser	Pro	Val 30	Ala	Ala
10	Ile	Val	Thr 35	Asp	Ile	Phe	Asn	Thr 40	Ser	Asp	Gly	Gly	Arg 45	Phe	Lys	Phe
	Pro	Asp 50	Gly	Val	Gln	Asn	Trp 55	Pro	Ala	Leu	Ser	Ile 60	Val	Ile	Ile	Ile
	Ile 65	Het	Thr	Ile	Gly	Gly 70	Asn	Ile	Leu	Val	Ile 75	Met	Ala	Val	Ser	Met 80
15	Glu	Lys	Lys	Leu	His 85	Asn	Ala	Thr	Asn	Tyr 90	Phe	Leu	Met	Ser	Leu 95	Ala
	Ile	Ala	Asp	Met 100	Leu	Val	Gly	Leu	Leu 105	Val	Met	Pro	Leu	Ser 110	Leu	Leu
20	Ala	Ile	Leu 115	туr	Asp	Tyr	Val	Trp 120	Pro	Leu	Pro	Arg	Tyr 125	Leu	Cys	Pro
	Val	Trp 130	Ile	Ser	Leu	Asp	Val 135	Leu	Phe	Ser	Thr	Ala 140	Ser	Ile	Met	His
	Leu 145	Cys	Ala	Ile	Ser	Leu 150	Asp	Arg	Tyr	Val	Ala 155	Ile	Arg	Asn	Pro	Ile 160
25	Glu	His	Ser	Arg	Phe 165	Asn	Ser	Arg	Thr	Lys 170	Ala	Ile	Met	Lys	Ile 175	Ala
	Ile	Val	Trp	Ala 180	Ile	Ser	Ile	Gly	Val 185	Ser	Val	Pro	Ile	Pro 190	Val	Ile
30	Gly	Leu	Arg 195	Asp	Glu	Glu	Lys	Val 200	Phe	Val	Asn	Asn	Thr 205	Thr	Cys	Val
	Leu	Asn 210	Asp	Pro	Asn	Phe	Val 215	Leu	Ile	Gly	Ser	Phe 220	Val	Ala	Phe	Phe
	Ile 225	Pro	Leu	Thr	Ile	Met 230	Val	Ile	Thr	Tyr	Cys 235	Leu	Thr	Ile	Tyr	Val 240
35	Leu	Arg	Arg	Gln	Ala	Leu	Met	Leu	Leu	His	Gly	His	Thr	Glu	Glu	Pro

250

255

	Pro	Gly	Leu	Ser 260	Leu	Asp	Phe	Leu	Lys 265	Cys	Cys	Lys	Arg	Asn 270	Thr	Ala	
	Glu	Glu	Glu 275	Asn	Ser	Ala	Asn	Pro 280	Asn	Glņ	Asp	Gln	Asn 285	Ala	Arg	Arg	
5	Arg	Lys 290	Lys	Lys	Glu	Arg	Arg 295	Pro	Arg	Gly	Thr	Met 300	Gln	Ala	Ile	Asn	
	Asn 305	Glu	Arg	Lys	Ala	Ser 310	Lys	Val	Leu	Gly	Ile 315	Val	Phe	Phe	Val	Phe 320	
10	Leu	Ile	Met	Trp	Cys 325	Pro	Phe	Phe	Ile	Thr 330	Asn	Ile	Leu	Ser	Val 335	Leu	
	Cys	Glu	Lys	Ser 340	Cys	Asn	Gln	Lys	Leu 345	Met	Glu	Lys	Leu	Leu 350	Asn	Val	
	Phe	Val	Trp 355	Ile	Gly	Tyr	Val	Cys 360	Ser	Gly	Ile	Asn	Pro 365	Leu	Val	Tyr	
15	Thr	Leu 370	Phe	Asn	Lys		Tyr 375	Arg	Arg	Ala		Ser 380	Asn	Tyr	Leu	Arg	
	Cys 385	Asn	Tyr	Lys	Val	Glu 390	Lys	Lys	Pro	Pro	Val. 395	Arg	Gln	Ile	Pro	Arg 400	
20	Val	Ala	Ala	Thr	Ala 405	Leu	Ser	Gly	Arg	Glu 410	Leu	Asn	Val	Asn ·	Ile 415	Tyr	
,		His		420					425	-			•	430			•
		Ile	435					440				Pro	Val 445	Asn	Pro	Ser	
25	•	Val 450	٠	÷			455		Ser	.Val		,				•	
	(128) IN	FORMI			-												
30		(B)	TYI STI	NGTH: PE: r RANDE	nucle EDNES	eic a SS: s	cid singl					•					
	(i	í) M		DLE 1				nomi	.c)	ini.	No Fee	∂n, 1 ° •					
	· (x.	i) SI	EQUE	NCE I	DESCR	RIPTI	ON:	SEQ	ID N	10:12	:7:	•		•			
35	GGTAAGCT						CTTC									ī	30
	(129) TN	FODM	ስጥተርነ	T EOE	O C C	תד ר	NO. 1	20.									

(129) INFORMATION FOR SEQ ID NO:128:

5	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:	
	TCCGAATTCT CTGTAGACAC AAGGCTTTGG	30
	(130) INFORMATION FOR SEQ ID NO:129:	
10	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1068 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
15	(ii) MOLECULE TYPE: DNA (genomic)	iğ.
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:	
	ATGGATCAGT TCCCTGAATC AGTGACAGAA AACTTTGAGT ACGATGATTT GGCTGAGGCC	60
	TGTTATATTG GGGACATCGT GGTCTTTGGG ACTGTGTTCC TGTCCATATT CTACTCCGTC	120
	ATCTTTGCCA TTGGCCTGGT GGGAAATTTG TTGGTAGTGT TTGCCCTCAC CAACAGCAAG	180
20	AAGCCCAAGA GTGTCACCGA CATTTACCTC CTGAACCTGG CCTTGTCTGA TCTGCTGTTT	240
	GTAGCCACTT TGCCCTTCTG GACTCACTAT TTGATAAATG AAAAGGGCCT CCACAATGCC	300
	ATGTGCAAAT TCACTACCGC CTTCTTCTTC ATCGGCTTTT TTGGAAGCAT ATTCTTCATC	360
	ACCGTCATCA GCATTGATAG GTACCTGGCC ATCGTCCTGG CCGCCAACTC CATGAACAAC	420
	CGGACCGTGC AGCATGGCGT CACCATCAGC CTAGGCGTCT GGGCAGCAGC CATTTTGGTG	480
25	GCAGCACCCC AGTTCATGTT CACAAAGCAG AAAGAAAATG AATGCCTTGG TGACTACCCC	540
	GAGGTCCTCC AGGAAATCTG GCCCGTGCTC CGCAATGTGG AAACAAATTT TCTTGGCTTC	600
	CTACTCCCCC TGCTCATTAT GAGTTATTGC TACTTCAGAA TCATCCAGAC GCTGTTTTCC	660
	TGCAAGAACC ACAAGAAAGC CAAAGCCATT AAACTGATCC TTCTGGTGGT CATCGTGTTT	720
	TTCCTCTTCT GGACACCCTA CAACGTTATG ATTTTCCTGG AGACGCTTAA GCTCTATGAC	780
30	TTCTTTCCCA GTTGTGACAT GAGGAAGGAT CTGAGGCTGG CCCTCAGTGT GACTGAGACG	840
	GTTGCATTTA GCCATTGTTG CCTGAATCCT CTCATCTATG CATTTGCTGG GGAGAAGTTC	900
	AGAAGATACC TTTACCACCT GTATGGGAAA TGCCTCCCTC TCCTGTCTCC CCCCTCACTC	060

CACGTTGATT TCTCCTCATC TGAATCACAA AGGAGCAGGC ATGGAAGTGT TCTGAGCAGC 1020

	CACGI	I GA			ı ÇAI	C 1G.	nn i C	ncan	AGG	MUCA	GGC ·	AIGG	MMG I	G1 1	CIGA	GCAG	C 1020
	AATTT	TAC	TT A	CCAC	ACGA	G TG	ATGG	AGAT	GCA'	TTGC	TCC	TTCT	CTGA				1068
	(131)	IN	FORM	ATIO	N FOI	R SE	Q ID	NO:	130:								
5		(i)	(A) (B) (C)) LEI) TYI) STI	CE CI NGTH PE: 8 RANDI POLO	: 35! amind EDNE:	5 am: o ac: SS:	ino a	acid	s		,	• .		e in Light		
		(ii) MC	OLECT	ULE :	TYPE	: pro	otei	n								
10		(xi	.) SI	EQUEI	NCE I	DESCI	RIPT:	ON:	SEQ	ID 1	NO:1	30:					
		Met 1	Asp	Gln	Phe	Pro 5	Glu	Ser	Val	Thr	Glu 10	Asn	Phe	Glu	Tyr	Asp	Asp
	:	Leu	Ala	Glu	Ala 20	Cys	Tyr	Ile	Gly	Asp 25	Ile	Val	Val	Phe	Gly 30	Thr	Val
15		Phe	Leu	Ser 35	Ile	Phe	Tyr	Ser	Val 40	Ile	Phe	Ala	Ile	Gly 45	Leu	Val	Gly
		Asn	Leu 50	Leu	Val	Val	Phe	Ala 55	Leu	Thr	Asn	Ser	Lys. 60	Lys	Pro	Lys	Ser
20	•	65					70		•			75		Asp			80
						85					90			Asn		95	_
					100			·		105	, :		; .	Phe	110		
25				115					120				•	Ile 125			
			130				-	135	•				140	Arg			
30	1	145					150					155		Ala	•		160
						165					170			Asn		175	
25					180					185				Val	190		
35	7	Val	Glu	Thr 195	Asn	Phe	Leu	Gly	Phe 200	Leu	Leu	Pro	Leu	Leu 205	Ile	Met	Ser

BNSDOCID <WO 002212941 1 3

		Cys Tyr 210	Phe	Arg	Ile	Ile 215	Gln	Thr	Leu	Phe	Ser 220	Cys	Lys	Asn	His	
	Lys : 225	Lys Ala	Lys		Ile 230	Lys	Leu	Ile	Leu	Leu 235	Val	Val	Ile	Val	Phe 240	
5	Phe 1	Leu Phe		Thr : 245	Pro				Met 250		Phe	Leu	Glu	Thr 255	Leu	
	Lys 1	Leu Tyr	Asp 260	Phe :	Phe	Pro	Ser	Cys 265	Asp	Met	Arg	Lys	Asp 270	Leu	Arg	
10	Leu <i>l</i>	Ala Leu 275	Ser	Val '	Thr		Thr 280	Val	Ala	Phe	Ser	His 285	Cys	Cys	Leu	
		Pro Leu 290	Ile	Tyr :		Phe 295	Ala	Gly	Glu	Lys	Phe 300	Arg	Arg	Tyr	Leu	
	Tyr I 305	His Leu	Tyr		Lys 310	Cys	Leu	Ala	Val	Leu 315	Cys	Gly	Arg	Ser	Val 320	
15	His V	Val Asp		Ser : 325	Ser	Ser	Glu	Ser	Gln 330	Arg	Ser	Arg	His	Gly 335	Ser	
	Val I	Leu Ser	Ser . 340	Asn i	Phe	Thr	Tyr	His 345	Thr	Ser	Asp	Gly	Asp 350	Ala	Leu	
20	Leu I	Leu Leu 355				•					•					
	(132) INFO	ORMATIO	1 FOR	SEQ	ID	NO:1	31:									
25	(i)	SEQUENC (A) LEI (B) TYI (C) STI (D) TOI	NGTH: PE: n: RANDE	32 l ucle: DNES!	base ic a S: s	pai cid ingl	rs									
	(ii)) MOLECT	JLE T	YPE:	DNA	(ge	nomi	.c)								
	(xi)) SEQUE	CE D	ESCR:	IPTI	ON:	SEQ	ID N	10:13	1:						
	GATCTCCAGT	r aggcar	TAAGT	GGA	CAAT	TCT	GG									32
30	(133) INFO	ORMATIO	1 FOR	SEQ	ID	NO:1										
35	(i)	SEQUENCE (A) LET (B) TYI (C) STI (D) TOI	NGTH: PE: n RANDE	30 l ucle: DNES:	base ic a S: s	pai cid ingl	rs			٠						
	(ii)) MOLECI	JLE T	YPE:	DNA	(ge	nomi	.c)								
	(xi)) SEQUE	ICE D	ESCR:	IPTI	ON:	SEQ	ID N	10:13	2:						

	CTCCTTCGGT CCTCCTATCG TTGTCAGAAG	30
	(134) INFORMATION FOR SEQ ID NO:133:	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:	
10	AGAAGGCCAA GATCGCGCGG CTGGCCCTCA	30
٠	(135) INFORMATION FOR SEQ ID NO:134:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:134:	
	CGGCGCCACC GCACGAAAAA GCTCATCTTC	30
20	(136) INFORMATION FOR SEQ ID NO:135:	
25	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:135: GCCAAGAAGC GGGTGAAGTT CCTGGTGGTG GCA	33
	(137) INFORMATION FOR SEQ ID NO:136:	
30	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
35	(ii) MOLECULE TYPE: DNA (genomic)	

	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:	
	CAGGCGGAAG GTGAAAGTCC TGGTCCTCGT	30
	(138) INFORMATION FOR SEQ ID NO:137:	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
10	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:	
	CGGCGCCTGC GGGCCAAGCG GCTGGTGGTG GTG	33
	(139) INFORMATION FOR SEQ ID NO:138:	
15	(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 31 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:138:	
20	CCAAGCACAA AGCCAAGAAA GTGACCATCA C	31
	(140) INFORMATION FOR SEQ ID NO:139:	
25	(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 30 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:139:	
	GCGCCGGCGC ACCAAATGCT TGCTGGTGGT	30
30	(141) INFORMATION FOR SEQ ID NO:140:	
•	(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 41 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single	
35	(D) TOPOLOGY: linear	

	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:	
	CAAAAAGCTG AAGAAATCTA AGAAGATCAT CTTTATTGTC G	4:
	(142) INFORMATION FOR SEQ ID NO:141:	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
10	(ii) MOLECULE TYPE: DNA (genomic)	•
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:141:	
	CAAGACCAAG GCAAAACGCA TGATCGCCAT	30
	(143) INFORMATION FOR SEQ ID NO:142:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:142:	
	GTCAAGGAGA AGTCCAAAAG GATCATCATC	30
	(144) INFORMATION FOR SEQ ID NO:143:	
25	(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 30 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:143:	
30	CGCCGCGTGC GGGCCAAGCA GCTCCTGCTC	30
•	(145) INFORMATION FOR SEQ ID NO:144:	
35	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	

PCT/US99/23938

		(ii) MOLECULE TYPE: DNA (genomic)	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO:144:	
	CCTGA	TAAGC GCTATAAAAT GGTCCTGTTT CGA	33
	(146)	INFORMATION FOR SEQ ID NO:145:	
5		(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 36 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
10		(ii) MOLECULE TYPE: DNA (genomic)	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO:145:	
	GAAAG	ACAAA AGAGAGTCAA GAGGATGTCT TTATTG	36
	(147)	INFORMATION FOR SEQ ID NO:146:	
15		 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	ر ان انگر
		(ii) MOLECULE TYPE: DNA (genomic)	\$t
20		(xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:	,
	CGGAG	AAAGA GGGTGAAACG CACAGCCATC GCC	33
	(148)	INFORMATION FOR SEQ ID NO:147:	
25		 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
		(ii) MOLECULE TYPE: DNA (genomic)	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO:147:	
30	AAGCT"	TCAGC GGGCCAAGGC ACTGGTCACC	30
	(149)	INFORMATION FOR SEQ ID NO:148:	
35		 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	

	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:148:	
	CAGCGGCAGA AGGCAAAAAG GGTGGCCATC	3(
	(150) INFORMATION FOR SEQ ID NO:149:	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
10	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:149:	
	CGGCAGAAGG CGAAGCGCAT GATCCTCGCG	30
	(151) INFORMATION FOR SEQ ID NO:150:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:150:	
	GAGCGCAACA AGGCCAAAAA GGTGATCATC	30
	(152) INFORMATION FOR SEQ ID NO:151:	
25	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:151:	
30	GGTGTAAACA AAAAGGCTAA AAACACAATT ATTCTTATT	39
	(153) INFORMATION FOR SEQ ID NO:152:	
35	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 27 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single	

PCT/US99/23938

	(ii) MOLECULE TYPE: DNA (genomic)			
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:152:		::	
	GAGAGCCAGC TCAAGAGCAC CGTGGTG			27
	(154) INFORMATION FOR SEQ ID NO:153:		• • .	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 			
10	(ii) MOLECULE TYPE: DNA (genomic)			
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:153:	٠		
	CCACAAGCAA ACCAAGAAAA TGCTGGCTGT			30
	(155) INFORMATION FOR SEQ ID NO:154:			
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 			
	(ii) MOLECULE TYPE: DNA (genomic)	,		
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:154:	ţ		
	CATCAAGTGT ATCATGTGCC AAGTACGCCC	٠		30
	(156) INFORMATION FOR SEQ ID NO:155:			
25	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 34 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 		÷	
	(ii) MOLECULE TYPE: DNA (genomic)			
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:155:			
30	CTAGAGAGTC AGATGAAGTG TACAGTAGTG GCAC			34
	(157) INFORMATION FOR SEQ ID NO:156:			
35	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 36 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 			

	(ii) MOLECULE TYPE: DNA (genomic)
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:156:
· :	CGGACAAAAG TGAAAACTAA AAAGATGTTC CTCATT 36
	(158) INFORMATION FOR SEQ ID NO:157:
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear
10	(ii) MOLECULE TYPE: DNA (genomic)
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:157:
	GCTGAGGTTC GCAATAAACT AACCATGTTT GTG
	(159) INFORMATION FOR SEQ ID NO:158:
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 29 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear
	(ii) MOLECULE TYPE: DNA (genomic)
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:158:
	GGGAGGCCGA GCTGAAAGCC ACCCTGCTC 29
	(160) INFORMATION FOR SEQ ID NO:159:
25	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear
	(ii) MOLECULE TYPE: DNA (genomic)
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:159:
30	CAAGATCAAG AGAGCCAAAA CCTTCATCAT G 31
	(161) INFORMATION FOR SEQ ID NO:160:
35	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear

BNSDOCID <WO 002212941 I >

	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:160:	
	CCGGAGACAA GTGAAGAAGA TGCTGTTTGT C	3]
	(162) INFORMATION FOR SEQ ID NO:161:	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
10	(ii) MOLECULE TYPE: DNA (genomic)	
	(x1) SEQUENCE DESCRIPTION: SEQ ID NO:161:	
	GCAAGGACCA GATCAAGCGG CTGGTGCTCA	30
	(163) INFORMATION FOR SEQ ID NO:162:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 34 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	,
	(ii) MOLECULE TYPE: DNA (genomic)	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:162:	
	CAAGAAAGCC AAAGCCAAGA AACTGATCCT TCTG	34
	(164) INFORMATION FOR SEQ ID NO:163:	
25	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1068 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:163:	
30	ATGGAAGATT TGGAGGAAAC ATTATTTGAA GAATTTGAAA ACTATTCCTA TGACCTAGAC	60
	TATTACTCTC TGGAGTCTGA TTTGGAGGAG AAAGTCCAGC TGGGAGTTGT TCACTGGGTC	120
	TCCCTGGTGT TATATTGTTT GGCTTTTGTT CTGGGAATTC CAGGAAATGC CATCGTCATT	180
	TGGTTCACGG GGCTCAAGTG GAAGAAGACA GTCACCACTC TGTGGTTCCT CAATCTAGCC	240
	ATTGCGGATT TCATTTTCT TCTCTTTCTG CCCCTGTACA TCTCCTATGT GGCCATGAAT	300

TTCCACTGGC CCTTTGGCAT CTGGCTGTGC AAAGCCAATT CCTTCACTGC CCAGTTGAAC 360

	ATGTTTGCCA GTGTTTTTT CCTGACAGTG ATCAGCCTGG ACCACTATAT CCACTTGATC 4	20
	CATCCTGTCT TATCTCATCG GCATCGAACC CTCAAGAACT CTCTGATTGT CATTATATTC 4	80
	ATCTGGCTTT TGGCTTCTCT AATTGGCGGT CCTGCCCTGT ACTTCCGGGA CACTGTGGAG 5	40
5	TTCAATAATC ATACTCTTTG CTATAACAAT TTTCAGAAGC ATGATCCTGA CCTCACTTTG 6	00
	ATCAGGCACC ATGTTCTGAC TTGGGTGAAA TTTATCATTG GCTATCTCTT CCCTTTGCTA 60	60
	ACAATGAGTA TTTGCTACTT GTGTCTCATC TTCAAGGTGA AGAAGCGAAC AGTCCTGATC 73	20
	TCCAGTAGGC ATAAGTGGAC AATTCTGGTT GTGGTTGTGG CCTTTGTGGT TTGCTGGACT 76	80
	CCTTATCACC TGTTTAGCAT TTGGGAGCTC ACCATTCACC ACAATAGCTA TTCCCACCAT 84	10
10	GTGATGCAGG CTGGAATCCC CCTCTCCACT GGTTTGGCAT TCCTCAATAG TTGCTTGAAC 90	00
	CCCATCCTTT ATGTCCTAAT TAGTAAGAAG TTCCAAGCTC GCTTCCGGTC CTCAGTTGCT 96	50
	GAGATACTCA AGTACACACT GTGGGAAGTC AGCTGTTCTG GCACAGTGAG TGAACAGCTC 102	20
	AGGAACTCAG AAACCAAGAA TCTGTGTCTC CTGGAAACAG CTCAATAA 106	8
	(165) INFORMATION FOR SEQ ID NO:164:	
15	(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 355 amino acids(B) TYPE: amino acid(C) STRANDEDNESS:(D) TOPOLOGY: not relevant	
20	(ii) MOLECULE TYPE: protein	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:164:	
	Met Glu Asp Leu Glu Glu Thr Leu Phe Glu Glu Phe Glu Asn Tyr Ser 1 10 15	
25	Tyr Asp Leu Asp Tyr Tyr Ser Leu Glu Ser Asp Leu Glu Glu Lys Val 20 25 30	
	Gln Leu Gly Val Val His Trp Val Ser Leu Val Leu Tyr Cys Leu Ala 35 40 45	
	Phe Val Leu Gly Ile Pro Gly Asn Ala Ile Val Ile Trp Phe Thr Gly 50 55 60	
30	Leu Lys Trp Lys Lys Thr Val Thr Thr Leu Trp Phe Leu Asn Leu Ala 65 70 75 80	
	Ile Ala Asp Phe Ile Phe Leu Leu Phe Leu Pro Leu Tyr Ile Ser Tyr 85 90 95	

	 Val	Ala	Met	Asn 100	Phe	Hiś	Trp	Pro	Phe 105	Gly	Ile	Trp	Leu	Cys 110	Lys	Ala
	Asn	Ser	Phe 115	Thr	Ala	Gln	Leu	Asn 120	Met	Phe	Ala	Ser	Val 125	Phe	Phe	Leu
5	Thr	Val 130	Ile	Ser	Leu	Asp	His 135	Tyr	Ile	His	Leu	11e 140	His	Pro	Val	Leu
	Ser 145	His	Arg	His	Arg	Thr 150	Leu	Lys	Asn	Ser	Leu 155	Ile	Val	Ile	Ile	Phe 160
10	Ile	Trp	Leu	Ľeu	Ala 165	Ser	Leu	Ile	Gly	Gly 170	Pro	Ala	Leu	Tyr	Phe 175	Arg
	Asp	Thr	Val	Glu 180	Phe	Asn	Asn	His	Thr 185	Leu	Cys	Tyr	Asn	Asn 190	Phe	Gln
	Lys	His	Asp 195	Pro	Asp	Leu	Thr	Leu 200	Ile	Arg	His	His	Val 205	Leu	Thr	Trp
15	Val	Lys 210	Phe	Ile	Ile	Gly	Tyr 215	Leu	Phe	Pro	Leu	Leu 220	Thr	Met	Ser	Ile
	Cys 225	Tyr	Leu	Cys	Leu	Ile 230	Phe	Lys	Val	Lys	Lys 235	Arg	Thr	Val	Leu	Ile 240
20	Ser	Ser	Arg	His	Lys 245	Trp	Thr	Ile	Leu	Val 250	Val	Val	Val	Ala	Phe 255	Val
	Val	Cys	Trp	Thr 260	Pro	Tyr	His	Leu	Phe 265	Ser	Ile	Trp	Glu	Leu 270	Thr	Ile
	His	His	Asn 275	Ser	Tyr	Ser	His	His 280	Val	Met	Gln	Ala	Gly 285	Ile	Pro	Leu
25	Ser	Thr 290		Leu	Ala	Phe	Leu 295	Asn	Ser	Cys	Leu	Asn 300	Pro	Ile	Leu	Tyr
	Val 305		Ile	Ser	Lys	Lys 310	Phe	Gln	Ala	Arg	Phe 315		Ser	Ser	Val	Ala 320
30	Glu	Ile	Leu	Lys	Туг 325		Leu	Trp	Glu	Val 330	Ser	Cys	Ser	Gly	Thr 335	Val
	Ser	Glu	Gln	Leu 340		Asn	Ser	Glu	Thr 345		Asn	Leu	Cys	Leu 350		Glu-
	Thr	Ala	Gln													

35 (166) INFORMATION FOR SEQ ID NO:165:

355

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1089 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- 5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:165:

ATGGGCAACC ACACGTGGGA GGGCTGCCAC GTGGACTCGC GCGTGGACCA CCTCTTTCCG CCATCCCTCT ACATCTTTGT CATCGGCGTG GGGCTGCCCA CCAACTGCCT GGCTCTGTGG 120 GCGGCCTACC GCCAGGTGCA ACAGCGCAAC GAGCTGGGCG TCTACCTGAT GAACCTCAGC 180 ATCGCCGACC TGCTGTACAT CTGCACGCTG CCGCTGTGGG TGGACTACTT CCTGCACCAC 240 10 GACAACTGGA TCCACGGCCC CGGGTCCTGC AAGCTCTTTG GGTTCATCTT CTACACCAAT 300 ATCTACATCA GCATCGCCTT CCTGTGCTGC ATCTCGGTGG ACCGCTACCT GGCTGTGGCC 360 CACCCACTCC GCTTCGCCCG CCTGCGCCGC GTCAAGACCG CCGTGGCCGT GAGCTCCGTG 420 GTCTGGGCCA CGGAGCTGGG CGCCAACTCG GCGCCCCTGT TCCATGACGA GCTCTTCCGA 480 GACCGCTACA ACCACACTT CTGCTTTGAG AAGTTCCCCA TGGAAGGCTG GGTGGCCTGG 15 ATGAACCTCT ATCGGGTGTT CGTGGGCTTC CTCTTCCCGT GGGCGCTCAT GCTGCTGTCG 600 TACCGGGGCA TCCTGCGGGC CGTGCGGGGC AGCGTGTCCA CCGAGCGCCA GGAGAAGGCC 660 AAGATCGCGC GGCTGGCCCT CAGCCTCATC GCCATCGTGC TGGTCTGCTT TGCGCCCTAT 720 CACGTGCTCT TGCTGTCCCG CAGCGCCATC TACCTGGGCC GCCCCTGGGA CTGCGGCTTC 780 GAGGAGCGCG TCTTTTCTGC ATACCACAGC TCACTGGCTT TCACCAGCCT CAACTGTGTG 20 GCGGACCCCA TCCTCTACTG CCTGGTCAAC GAGGGCGCCC GCAGCGATGT GGCCAAGGCC 900 CTGCACAACC TGCTCCGCTT TCTGGCCAGC GACAAGCCCC AGGAGATGGC CAATGCCTCG 960 CTCACCCTGG AGACCCCACT CACCTCCAAG AGGAACAGCA CAGCCAAAGC CATGACTGGC 1020 AGCTGGGCGG CCACTCCGCC TTCCCAGGGG GACCAGGTGC AGCTGAAGAT GCTGCCGCCA 1080 GCACAATGA 1089

- 25 (167) INFORMATION FOR SEQ ID NO:166:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 362 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
- 30 (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein

	(xi) SEQUENCE DESCRIPTION:						SEQ	ID 1	10:16	56:								
		Met 1	Gly	Asn	His	Thr 5	Trp	Glu	Gly	Cys	His 10	Val	Asp	Ser	Arg	Val 15	Asp	
5		His	Leu	Phe	Pro 20	Pro	Ser	Leu	Tyr	Ile 25	Phe	Val.	Ile	Gly	Val 30	Gly	Leu	
		Pro	Thr	Asn 35	Cys	Leu	Ala	Leu	Trp 40	Ala	Ala	Tyr	Arg	Gln 45	Val	Gln	Gln	
	• .	Arg	Asn 50	Glu	Leu	Gly	Val	Tyr 55	Leu	Met	Asn	Leu	Ser 60	Ile	Ala	Asp	Leu	
10		Leu 65	туг	Ile	Cys	Thr	Leu 70	Pro	Leu	Trp	Val	Asp 75	Tyr	Phe	Leu	His	His 80	
	٠.	Asp	Asn	Trp	Ile	His 85	Gly	Pro	Gly	Ser	Cys 90	Lys	Leu	Phe	Gly	Phe 95	Ile	
15		Phe	туг	Thr	Asn 100	Ile	Tyr	Ile	Ser	Ile 105	Ala	Phe	Leu	Cys	Cys 110	Ile	Ser	
		Val	_	Arg	Tyr	Leu	Ala	Val	Ala 120	His	Pro	Leu	Arg	Phe 125	Ala	Arg	Leu	
	. •	Arg	Arg 130	Val	Lys	Thr	Ala	Val	Ala	Val	Ser	Ser	Val 140	Val	Trp	Ala	Thr	
20		Glu 145	Leu	Gly	Ala	Asn	Ser 150	Ala	Pro	Leu	Phe	His 155	Asp	Glu	Leu	Phe	Arg 160	
		Asp	Arg	Tyr	Asn	His 165	Thr	Phe	Cys	Phe	Glu 170	Lys	Phe	Pro	Met	Glu 175	Gly	
25		Trp	Val	Ala	Trp 180	Met	Asn	Leu	Tyr	Arg 185	Val	Phe	Val	Gly	Phe 190	Leu	Phe	
		Pro	Trp	Ala 195	Leu	Met	Leu	Leu	Ser 200	Tyr	Arg	Gly	Ile	Leu 205	Arg	Ala	Val	
		Arg	Gly 210	Ser	Val	Ser	Thr	Glu 215	Arg	Gln	Glu	Lys	Ala 220	Lys	Ile	Ala	Arg	
30		Leu 225		Leu	Ser	Leu	Ile 230		Ile	Val	Leu	Val 235	Cys	Phe	Ala	Pro	Tyr 240	
		His	Val	Leu	Leu	Leu 245		Arg	Ser	Ala	Ile 250	туr	Leu	Gly	Arg	Pro 255	Trp	
35		Asp	Cys	Gly	Phe 260		Glu	Arg	Val	Phe 265		Ala	Tyr	His	Ser 270	Ser	Leu	
		Ala	Phe	Thr	Ser	Leu	Asn	Cys	Val	Ala	Asp	Pro	Ile	Leu	Tyr	Cys	Leu	

	275 280 2	85
	Val Asn Glu Gly Ala Arg Ser Asp Val Ala Lys Ala Lo 290 295 300	eu His Asn Leu
5	Leu Arg Phe Leu Ala Ser Asp Lys Pro Gln Glu Met A 5 305 310 315	la Asn Ala Ser 320
	Leu Thr Leu Glu Thr Pro Leu Thr Ser Lys Arg Asn Se	er Thr Ala Lys 335
	Ala Met Thr Gly Ser Trp Ala Ala Thr Pro Pro Ser Gl	ln Gly Asp Gln
10	O Val Gln Leu Lys Met Leu Pro Pro Ala Gln 355 360	
	(168) INFORMATION FOR SEQ ID NO:167:	, ·
15	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1002 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	•
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:167:	
20	0 ATGGAGTCCT CAGGCAACCC AGAGAGCACC ACCTTTTTT ACTATGACCT	TCAGAGCCAG 60
	CCGTGTGAGA ACCAGGCCTG GGTCTTTGCT ACCCTCGCCA CCACTGTCCT	GTACTGCCTG 120
	GTGTTTCTCC TCAGCCTAGT GGGCAACAGC CTGGTCCTGT GGGTCCTGGT	GAAGTATGAG 180
	AGCCTGGAGT CCCTCACCAA CATCTTCATC CTCAACCTGT GCCTCTCAGA	CCTGGTGTTC 240
	GCCTGCTTGT TGCCTGTGTG GATCTCCCCA TACCACTGGG GCTGGGTGCT	GGGAGACTTC 300
25	5 CTCTGCAAAC TCCTCAATAT GATCTTCTCC ATCAGCCTCT ACAGCAGCAT	CTTCTTCCTG 360
	ACCATCATGA CCATCCACCG CTACCTGTCG GTAGTGAGCC CCCTCTCCAC	CCTGCGCGTC 420
	CCCACCCTCC GCTGCCGGGT GCTGGTGACC ATGGCTGTGT GGGTAGCCAG	CATCCTGTCC 480
	TCCATCCTCG ACACCATCTT CCACAAGGTG CTTTCTTCGG GCTGTGATTA	TTCCGAACTC 540
	ACGTGGTACC TCACCTCCGT CTACCAGCAC AACCTCTTCT TCCTGCTGTC	CCTGGGGATT 600
30	ATCCTGTTCT GCTACGTGGA GATCCTCAGG ACCCTGTTCC GCTCACGCTC	CAAGCGGCGC 660
	CACCGCACGA AAAAGCTCAT CTTCGCCATC GTGGTGGCCT ACTTCCTCAG	CTGGGGTCCC 720
	TACAACTTCA CCCTGTTTCT GCAGACGCTG TTTCGGACCC AGATCATCCG	GAGCTGCGAG 780

.....

117

GCCAAACAGC AGCTAGAATA CGCCCTGCTC ATCTGCCGCA ACCTCGCCTT CTCCCACTGC 840 TGCTTTAACC CGGTGCTCTA TGTCTTCGTG GGGGTCAAGT TCCGCACACA CCTGAAACAT 900 GTTCTCCGGC AGTTCTGGTT CTGCCGGCTG CAGGCACCCA GCCCAGCCTC GATCCCCCAC 960 TCCCCTGGTG CCTTCGCCTA TGAGGGCGCC TCCTTCTACT GA 1002 (169) INFORMATION FOR SEQ ID NO:168: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 333 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: 10 (D) TOPOLOGY: not relevant (ii) MOLECULE TYPE: protein (xi) SEQUENCE DESCRIPTION: SEQ ID NO:168: Met Glu Ser Ser Gly Asn Pro Glu Ser Thr Thr Phe Phe Tyr Tyr Asp 15 Leu Gln Ser Gln Pro Cys Glu Asn Gln Ala Trp Val Phe Ala Thr Leu 25 Ala Thr Thr Val Leu Tyr Cys Leu Val Phe Leu Leu Ser Leu Val Gly Asn Ser Leu Val Leu Trp Val Leu Val Lys Tyr Glu Ser Leu Glu Ser 💆 20 Leu Thr Asn Ile Phe Ile Leu Asn Leu Cys Leu Ser Asp Leu Val Phe 70 75 Ala Cys Leu Leu Pro Val Trp Ile Ser Pro Tyr His Trp Gly Trp Val 85 25 Leu Gly Asp Phe Leu Cys Lys Leu Leu Asn Met Ile Phe Ser Ile Ser 105 Leu Tyr Ser Ser Ile Phe Phe Leu Thr Ile Met Thr Ile His Arg Tyr 115 Leu Ser Val Val Ser Pro Leu Ser Thr Leu Arg Val Pro Thr Leu Arg 30 140 135 Cys Arg Val Leu Val Thr Met Ala Val Trp Val Ala Ser Ile Leu Ser 155 Ser Ile Leu Asp Thr Ile Phe His Lys Val Leu Ser Ser Gly Cys Asp 170 Tyr Ser Glu Leu Thr Trp Tyr Leu Thr Ser Val Tyr Gln His Asn Leu 35

185

	Phe		eu Le .95	ı Ser	Leu	Gly	Ile 200	Ile	Leu	Phe	Cys	Tyr- 205	Val	Glu	Ile	
		Arg T 210	hr Le	ı Phe	Arg	Ser 215	Arg	Ser	Lys	Arg	Arg 220	His	Arg	Thr	Lys	
5	Lys 225	Leu I	le Ph	e Ala	Ile 230	Val	Val	Ala	Tyr	Phe 235	Leu	Ser	Trp	Gly	Pro 240	
	Tyr	Asn P	he Th	r Leu 245	Phe	Leu	Gln	Thr	Leu 250	Phe	Arg	Thr	Gln	Ile 255	Ile	
10	Arg	Ser C	Cys Gl 26		Lys	Gln	Gln	Leu 265	Glu	Tyr	Ala	Leu	Leu 270	Île	Cys	
	Arg		Leu Al 275	a Phe	Ser	His	Cys 280	Cys	Phe	Asn	Pro	Val 285		Tyr	Val	
	Phe	Val 0 290	Gly Va	l Lys	Phe	Arg 295	Thr	His	Leu	Lys	His 300	Väl	Leu	Arg	Gln	
15	Phe 305	Trp F	Phe Cy	s Arg	Leu 310		Ala	Pro	Ser	Pro 315	Ala	Ser	Ile	Pro	His 320	
	Ser	Pro 0	Gly Al	a Phe 325		Tyr	Glu	Gly	Ala 330	Ser	Phe	Tyr				
	(170) IN	FORMA"	TION F	OR SE	Q ID	NO:	169:									
20	(i)	(A) (B) (C)	UENCE LENGT TYPE: STRAN TOPOI	H: 98 nucl DEDNE	87 ba eic ESS:	se p acid sing	airs	•								
25	(i:	i) MO	LECULE	турі	E: DN	IA (g	enom	ic)								
•	(x:	i) SE	QUENCI	DESC	: CRIPT	'ION:	SEQ	ID	NO:1	69:	•					
,	ATGGACAA	CG CC'	TCGTT	TC G	GAGCC	CTGG	CCC	GCCA	ACG	CATC	GGGC	CC G	GACC	CGGC	G	6
	CTGAGCTG	CT CC	AACGC	STC G	ACTCI	rggcg	CCG	CTGC	CGG	CGCC	GCTG	GC G	GTGG	CTGT	A	12
	CCAGTTGT	CT AC	GCGGT	SAT C	rgcgc	CGTG	GGT	CTGG	CGG	GCAA	CTCC	GC C	GTGC	TGTA	C	18
30	GTGTTGCT	GC GG	GCGCC	CCG C	ATGAZ	AGACO	GTC	ACCA	AACC	TGTI	CATO	CT C	CAACC	TGGC	С	24
	ATCGCCGA	CG AG	CTCTT	CAC G	CTGGT	rgcto	CCC	CATC	AACA	TCGC	CGAC	CTT C	CTGC	TGCG	G	30
	CAGTGGCC	CT TC	GGGGA	GCT C.	ATGT	CAAC	G CTC	CATCO	STGG	CTAT	CGAC	CA C	STACA	ACAC	C.	36
	TTCTCCAG	CC TC	TACTT	CCT C	ACCG7	CATO	AGC	CGCC	GACC	GCTA	CCTC	GT (GTGI	TGGC	C.C	42
	ACTGCGGA	GT CG	CGCCG	GGT G	GCCG	CCG	C ACC	CTAC	AGCG	CCGC	GCG	GC (GTG#	GCCI	'G	48

	GCCGTGTG	GG GG	SATCO	STCAC	ACT	CGT	CGTG	CTG	CCTT	rcg ·	CAGT	CTTC	GC C	cggc:	raga	2 !	540
	GACGAGCAG	GG GC	CCGG	CGCCA	GTC	CGT	GCTA	GTC	TTC	CGC 2	AGCC	CGAG	GC C	TTCT	GTG	3 6	500
	CGCGCGAG	CC GC	CTCT	racac	GC1	CGT	CTG	GGC	rtcg	CCA '	rccc	CGTG	rc cz	ACCA!	rctg:	г е	660
	GTCCTCTAT	ra co	CACCO	CTGCI	GTC	CCG	GCTG	CATO	CCA	rgc (GGCT	GAC	AG C	CACG	CCAA	3 7	720
5	GCCCTGGAG	GC GC	CGCCI	AAGAA	GCC	GGT	BAAG	TTC	TGGT	rgg '	TGGC2	AATC	CT GO	GCGG	rgtgo	7	780
	CTCCTCTG	CT GO	BACGO	CCTA	CCZ	ACCT	GAGC	ACC	TGGT	rgg (CGCT	CACC	AC CO	GACCI	rccc	3 8	340
	CAGACGCCC	GC TO	GTC	ATCGC	TAT	CTC	CTAC	TTC	ATCAC	CCA (GCCT	GACG	ra co	GCCA	ACAGO	2 9	900
	TGCCTCAAC	ee ee	CTTC	CTCTA	/ CGC	CTT	CCTG	GAC	CCAC	GCT '	rccg	CAGG	AA C	CTCC	GCCA(3 9	960
	CTGATAACT	rt Go	CCGCC	GCGGC	: AGC	CTG	A									9	987
10	(171) INI	FORM	OITA	1 FOF	SEÇ	O ID	NO:	L70:									
	(i)	(A) (B)	LEN	CE CH NGTH: PE: & RANDE	328 mino	am:	ino a		5								ઇ.
15		(D)	TOI	POLOG	5Y: r	ot i	relev	vant	•								
	(i:	i) MC	DLECT	JLE 1	YPE:	pro	oteir	מ									
	(x:	i) SE	EQUE	NCE I	ESCI	RIPT	ION:	SEQ	ID 1	10:1	70:						
	Met 1	Asp	Asn	Ala	Ser 5	Phe	Ser	Glu	Pro	Trp 10	Pro	Ala	Asn	Ala	Ser 15	Gly	٠
20	Pro	Asp	Pro	Ala 20	Leu	Ser	Cys	Ser	Asn 25	Ala	Ser	Thr	Leu	Ala 30	Pro	Leu	
	Pro	Ala	Pro 35	Leu	Ala	Val	Ala	Val 40	Pro	Val	Val	Tyr	Ala 45	Val	Ile	Cys	٠
25	Ala	Val 50	Gly	Leu	Ala	Gly	Asn 55	Ser	Ala	Val	Leu	Tyr 60	Val	Leu	Leu	Arg	
	Ala 65	Pro	Arg	Met	Lys	Thr 70	Val	Thr	Asn	Leu	Phe 75	Ile	Leu	Asn	Leu	Ala 80	
	Ile	Ala	Asp	Glu	Leu 85	Phe	Thr	Leu	Val	Leu 90	Pro	Ile	Asn	Ile	Ala 95	Asp	
30	Phe	Leu	Leu	Arg 100	Gln	Trp	Pro	Phe	Gly 105	Glu	Leu	Met	Cys	Lys 110	Leu	Ile	
	. Val	Ala	Ile 115	Asp	Gln	Tyr	Asn	Thr 120	Phe	Ser	Ser	Leu	Tyr 125	Phe	Leu	Thr	
	Val	Met	Ser	Ala	Asp	Arg	Tyr	Leu	Val	Val	Leu	Ala	Thr	Ala	Glu	Ser	

		٠.	130				•	135		٠.	1,		140			* .	
		Arg 145	Arg	Val	Ala	Gly	Arg 150	Thr	Tyr	Ser	Ala	Ala 155	Arg	Ala	Val	Ser	Leu 160
5	•	Ala	Val	Trp	Gly	Ile 165		Thr		Val	Val 170	Leu	Pro	Phe	Ala	Val 175	
		Ala	Arg	Leu	Asp 180	Asp	Glu	Gln	Gly	Arg 185	Arg	Gln	Cys	Val	Leu 190	Val	Phe
		Pro	Gln	Pro 195	Glu	Ala	Phe	Trp	Trp 200	_	Ala	Ser	Arg	Leu 205	-	Thr	Leu
10		Val	Leu 210	Gly	Phe	Ala	Ile	Pro 215	Val	Ser	Thr	Ile	Cys 220	Val	Leu	Tyr	Thr
		Thr 225	Leu	Leu	Cys	Arg				Met	_	Leu 235	Asp		His	Ala	Lys 240
15		Ala	Leu	Glu	Arg	Ala 245	Lys	Lys	Arg	Val	Lys 250	Phe	Leu	Val	Val	Ala 255	Ile
		Leu	Ala	Val	Cys 260	Leu	Leu	Cys	Trp	Thr 265		Tyr	His	Leu	Ser 270	Thr	Val
		Val	Ala	Leu 275	Thr	Thr	Asp	Leu	Pro 280	Gln	Thr	Pro	Leu	Val 285	Ile	Ala	Ile
20		Ser	Tyr 290	Phe	Ile	Thr	Ser	Leu 295	Thr	Tyr	Ala	Asn	Ser 300	Cys	Leu	Asn	Pro
		Phe 305	Leu	Tyr	Ala	Phe	Leu 310	Asp	Ala	Ser	Phe	Arg 315	Arg	Asn :	Leu	Arg	Gln 320
25		Leu	Ile	Thr	Cys	Arg 325	Ala	Ala	Ala						:		
	(172) IN	FORM	OITA	v FOI	R SE	Q ID	NO:	L71:								
		(i)	(A)	QUENC	IGTH:	: 100	02 ba	ase p		5			•				
30			(C)	STI	RANDI	EDNE	SS: 5	singl	le								
		(i:	i) Mo	OLEC	JLE :	TYPE	: DN	A (ge	enom:	ic)							
		(x:	i) S1	EQUE	NCE I	DESCI	RIPT:	ION:	SEQ	ID 1	10:1	71:	÷				
	ATGC	AGGC	CG C	rggg	CACC	C AG	AGCC	CCTT	GAC	AGCA	GG (CTC	CTTCT	rc co	CTCC	CAC	60
35	ATGG	GTGC	CA AC	CGTC	rctc/	A GG	ACAA!	rggc	ACTO	GCC2	ACA 2	ATGC	CACCI	TT CI	rccgi	AGCC	120
	CTGC	CGTT	CC TO	CTATO	GTGC	r cc	rgcc	CGCC	GTG:	CACT	CCG (GAT	CTGT	C TO	TGG	GCT	3 180

PCT/US99/23938

	ACTGGCAACA CGGCCGTCAT CCTTGTAATC CTAAGGGCGC CCAAGATGAA GACGGTGACC	240
	AACGTGTTCA TCCTGAACCT GGCCGTCGCC GACGGGCTCT TCACGCTGGT ACTGCCTGTC	300
	AACATCGCGG AGCACCTGCT GCAGTACTGG CCCTTCGGGG AGCTGCTCTG CAAGCTGGTG	360
	CTGGCCGTCG ACCACTACAA CATCTTCTCC AGCATCTACT TCCTAGCCGT GATGAGCGTG	420
5	GACCGATACC TGGTGGTGCT GGCCACCGTG AGGTCCCGCC ACATGCCCTG GCGCACCTAC	480
	CGGGGGGCGA AGGTCGCCAG CCTGTGTGTC TGGCTGGGCG TCACGGTCCT GGTTCTGCCC	540
	TTCTTCTCTT TCGCTGGCGT CTACAGCAAC GAGCTGCAGG TCCCAAGCTG TGGGCTGAGC	600
	TTCCCGTGGC CCGAGCAGGT CTGGTTCAAG GCCAGCCGTG TCTACACGTT GGTCCTGGGC	660
	TTCGTGCTGC CCGTGTGCAC CATCTGTGTG CTCTACACAG ACCTCCTGCG CAGGCTGCGG	720
10	GCCGTGCGGC TCCGCTCTGG AGCCAAGGCT CTAGGCAAGG CCAGGCGGAA GGTGAAAGTC	780
	CTGGTCCTCG TCGTGCTGGC CGTGTGCCTC CTCTGCTGGA CGCCCTTCCA CCTGGCCTCT	840
	GTCGTGGCCC TGACCACGGA CCTGCCCCAG ACCCCACTGG TCATCAGTAT GTCCTACGTC	900
	ATCACCAGCC TCACGTACGC CAACTCGTGC CTGAACCCCT TCCTCTACGC CTTTCTAGAT	960
	GACAACTTCC GGAAGAACTT CCGCAGCATA TTGCGGTGCT GA	002
15	(173) INFORMATION FOR SEQ ID NO:172:	•
20	(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 333 amino acids(B) TYPE: amino acid(C) STRANDEDNESS:	
20	(D) TOPOLOGY: not relevant	
	(ii) MOLECULE TYPE: protein	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:172:	
	Met Gln Ala Ala Gly His Pro Glu Pro Leu Asp Ser Arg Gly Ser Phe 1 5 10 15	
25	Ser Leu Pro Thr Met Gly Ala Asn Val Ser Gln Asp Asn Gly Thr Gly 20 25 30	٠
	His Asn Ala Thr Phe Ser Glu Pro Leu Pro Phe Leu Tyr Val Leu Leu 35 40 45	
30	Pro Ala Val Tyr Ser Gly Ile Cys Ala Val Gly Leu Thr Gly Asn Thr 50 55 60	
	Ala Val Ile Leu Val Ile Leu Arg Ala Pro Lys Met Lys Thr Val Thr	

65 70 75 80

		Asn	Val	Phe	Ile	Leu 85	Asn	Leu	Ala	Val	Ala 90	Asp	Gly	Leu	Phe	Thr 95	Leu	
		Val	Leu	Pro	Val 100	Asn	Ile	Ala	Glu	His 105	Leu	Leu	Gln	Tyr	Trp 110	Pro	Phe	
5		Gly	Glu	Leu 115	Leu	Cys	Lys	Leu	Val 120	Leu	Ala	Val	Asp	His 125	Tyr	Asn	Ile	
		Phe	Ser 130	Ser	Ile	Tyr	Phe	Leu 135	Ala	Val	Met	Ser	Val 140	Asp	Arg	Tyr	Leu	
10		Val 145	Val	Leu	Ala	Thr	Val 150	Arg	Ser	Arg	His	Met 155	Pro	Trp	Arg	Thr	Tyr 160	
		Arg	Gly	Ala	Lys	Val 165	Ala	Ser	Leu	Cys	Val 170	Trp	Leu	Gly	Val	Thr 175	Val	
•		Leu	Val	Leu	Pro 180	Phe	Phe	Ser	Phe	Ala 185	Gly	Val	Tyr	Ser	Asn 190	Glu	Leu	
15		Gln	Val	Pro 195	Ser	Cys	Gly	Leu	Ser 200	Phe	Pro	Trp	Pro	Glu 205	Gln	Val	Trp	
;		Phe	Lys 210	Ala	Ser	Arg	Val	Tyr 215	Thr	Leu	Val	Leu	Gly 220		Val	Leu	Pro	
20		Val 225	Cys	Thr	Ile	Cys	Val 230	Leu	Tyr	Thr	Asp	Leu 235	Leu	Arg	Arg	Leu	Arg 240	
		Ala	Val	Arg	Leu	Arg 245	Ser	Gly	Ala	Lys	Ala 250	Leu	Gly	Lys	Ala	Arg 255	Arg	
		Lys	Val	Lys	Val 260	Leu	Val	Leu	Val	Val 265	Leu	Ala	Val	Cys	Leu 270	Leu	Cys	
25		Trp	Thr	Pro 275	Phe	His	Leu	Ala	Ser 280		Val	Ala	Leu	Thr 285	Thr	Asp	Leu	
		Pro	Gln 290	Thr	Pro	Leu	Val	Ile 295	Ser	Met	Ser	Tyr	Val 300	Ile	Thr	Ser	Leu	
30		Thr 305	Tyr	Ala	Asn	Ser	Cys 310			Pro	Phe	Leu 315	Tyr	Ala	Phe	Leu	Asp 320	
		Asp	Asn	Phe	_	_		Phe	_		Ile 330	Leu	Arg	Cys	, •			
	(174) IN	FORM	OITA	N FO	R SE	Q ID	NO:	173:									
35		(i)	(A) LE	NGTH	: 11	07 b	ISTI ase acid	pair	s								

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

123

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:173:

to give the graph and the control of the control and the control of the control of the control of the control of

	ATGGTCCTTG	AGGTGAGTGA	CCACCAAGTG	CTAAATGACG	CCGAGGTTGC	CGCCCTCCTG	6
	GAGAACTTCA	GCTCTTCCTA	TGACTATGGA	GAAAACGAGA	GTGACTCGTG	CTGTACCTCC	12
5	CCGCCCTGCC	CACAGGACTT	CAGCCTGAAC	TTCGACCGGG	CCTTCCTGCC	AGCCCTCTAC	18
	AGCCTCCTCT	TTCTGCTGGG	GCTGCTGGGC	AACGGCGCGG	TGGCAGCCGT	GCTGCTGAGC	. 24
	CGGCGGACAG	CCCTGAGCAG	CACCGACACC	TTCCTGCTCC	ACCTAGCTGT	AGCAGACACG	30
	CTGCTGGTGC	TGACACTGCC	GCTCTGGGCA	GTGGACGCTG	CCGTCCAGTG	GGTCTTTGGC	36
	TCTGGCCTCT	GCAAAGTGGC	AGGTGCCCTC	TTCAACATCA	ACTTCTACGC	AGGAGCCCTC	420
10	CTGCTGGCCT	GCATCAGCTT	TGACCGCTAC	CTGAACATAG	TTCATGCCAC	CCAGCTCTAC	48
	CGCCGGGGGC	CCCCGGCCCG	CGTGACCCTC	ACCTGCCTGG	CTGTCTGGGG	GCTCTGCCTG	54
	CTTTTCGCCC	TCCCAGACTT	CATCTTCCTG	TCGGCCCACC	ACGACGAGCG	CCTCAACGCC	.60
	ACCCACTGCC	AATACAACTT	CCCACAGGTG	GGCCGCACGG	CTCTGCGGGT	GCTGCAGCTG	. 66
	GTGGCTGGCT	TTCTGCTGCC	CCTGCTGGTC	ATGGCCTACT	GCTATGCCCA	CATCCTGGCC	. 72
15	GTGCTGCTGG	TTTCCAGGGG	CCAGCGGCGC	CTGCGGGCCA	AGCGGCTGGT	GGTGGTGGTC	780
	GTGGTGGCCT	TTGCCCTCTG	CTGGACCCCC	TATCACCTGG	TGGTGCTGGT	GGACATCCTC	84
	ATGGACCTGG	GCGCTTTGGC	CCGCAACTGT	GGCCGAGAAA	GCAGGGTAGA	CGTGGCCAAG	900
	TCGGTCACCT	CAGGCCTGGG	CTACATGCAC	TGCTGCCTCA	ACCCGCTGCT	CTATGCCTTT	960
	GTAGGGGTCA	AGTTCCGGGA	GCGGATGTGG	ATGCTGCTCT	TGCGCCTGGG	CTGCCCCAAC	102
20	CAGAGAGGGC	TCCAGAGGCA	GCCATCGTCT	TCCCGCCGGG-	ATTCATCCTG	GTCTGAGACC	108
	TCAGAGGCCT	CCTACTCGGG	CTTGTGA				110

(175) INFORMATION FOR SEQ ID NO:174:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 368 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:

- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:174:

										12	24							
			Met 1	Val	Leu	Glu	Val 5	Ser	Asp	His	Gln	Val 10	Leu	Asn	Asp	Ala	Glu 15	Val
			Ala	Ala	Leu	Leu 20	Glu	Asn	Phe	Ser	Ser 25	Ser	Tyr	Asp	Tyr	Gly 30	Glu	Asn
	5		Glu	Ser	Asp 35	Ser	Cys	Cys	Thr	Ser 40	Pro	Pro	Cys	Pro	Gln 45	Asp	Phe	Ser
			Leu	Asn 50	Phe	Asp	Arg	Ala	Phe 55	Leu	Pro	Ala	Leu	Tyr 60	Ser	Leu	Leu	Phe
	10		Leu 65	Leu	Gly	Leu	Leu	Gly 70	Asn	Gly	Ala	Val	Ala 75	Ala	Val	Leu	Leu	Ser 80
			Arg	Arg	Thr	Ala	Leu 85	Ser	Ser	Thr	Asp	Thr 90	Phe	Leu	Leu	His	Leu 95	Ala
	*		Val	Ala	Asp	Thr 100	Leu	Leu	Val	Leu.	Thr 105	Leu	Pro	Leu	Trp	Ala 110	Val	Asp
	15		Ala	Ala	Val	Gln	Trp	Val	Phe	Gly 120	Ser	Gly	Leu.	Cys	Lys 125	Val	Ala	Gly
•			Ala	Leu 130	Phe	Asn	Ile		Phe 135	Tyr	Ala	Gly	Ala	Leu 140	Leu	Leų	Ala	Cys
, t.	20		Ile 145	Ser	Phe	Asp	Arg	Tyr 150	Leu	Asn	Ile	Val	His 155	Ala	Thr	Gln.	Leu	Tyr 160
:			Arg	Arg	Gly	Pro	Pro 165	Ala	Arg	Val	Thr	Leu 170	Thr	Cys	Leu	Ala	Val 175	
,		,	Gly	Leu	Cys	Leu 180	Leu	Phe	Ala	Leu	Pro 185	Asp	Phe	Ile	Phe	Leu 190	.Ser	Ala
	25		His	His	Asp 195	Glu	Arg	Leu	Asn	Ala 200	Thr	His	Cys	Gln	Tyr 205	Asn	Phe	Pro
			Gln	Val 210	Gly	Arg	Thr	Ala	Leu 215	Arg	Val	Leu	Gln	Leu 220	Val	Ala	Gly	Phe
	30		Leu 225	Leu	Pro	Leu	Leu	Val 230	Met	Ala	Tyr	Cys	Tyr 235	Ala	His	Ile	Leu	Ala 240
			Val	Leu	Leu	Val	Ser 245	Arg	Gly	Gln	Arg	Arg 250	Leu	Arg	Ala	Lys	Arg 255	Leu
			Val	Val	Val	Val 260		Val	Ala		Ala 265		Cys	Trp	Thr	Pro 270	Tyr	His
	35		Leu	Val	Val 275	Leu	Val	Asp	lle	Leu 280	Met	Asp	Leu	Gly	Ala 285	Leu	Ala	Arg
							~ 1		7			373	λ1 -	Tare	20-	Val	mb	

		90	*	295	• • •	•			300					
	Gly L 305	eu Gly Ty		is Cys 10	Cys	Leu	Asn	Pro 315	Leu	Leu	туr	Ala	Phe 320	
5		ly Val Ly						Met	Leu	Leu	Leu	Arg 335	Leu	
	Gly C	ys Pro As 34		rg Gly	Leu	Gln 345	Arg	Gln	Pro	Ser	Ser 350	Ser	Arg	
	Arg A	sp Ser Se 355	r Trp S	er Glu	Thr 360	Ser	Glu	Ala		Tyr 365	Ser	Gly	Leu	
10	(176) INFO	RMATION F	OR SEO	י אט.	175.									
		SEQUENCE												
		(A) LENGT (B) TYPE:	H: 1074	base	pairs	5								
15		(C) STRAN (D) TOPOL	DEDNESS	sing										
		MOLECULE			enomi	c)								
		SEQUENCE					0.17	· .						
	ATGGCTGATG								ידא ככ	תייחיים	እ ርጥጥ	ירא א רי	,	60,
	TTCACTGACT			*										
20	CCCTTGTACT		•											20
-0	TACTGGTACT													80
														40
	GCTGACCTCC													00
	AAGTTCCAGA													60
15	TGTGTGTTGC													20
25	AGAGCACATA													80
	TGGGTATTGG	CAGCTGCT	CT CTGC	TCCCA	GAAA	TCTT	A TA	CAGC	CAAA	T CA	AGGA	.GGAA	. 5	40
	TCCGGCATTG												_	00
	GCTGTCTTGA	CCCTGAAG	GT CATTO	TGGGG	TTCT	TCCT	rc c	CTTC	GTGG	T CA	TGGC	TTGC	: 6	60
	TGCTATACCA	TCATCATT	CA CACCO	TGATA	CAAG	CCAA	GA A	GTCT	TCCA	A GC	ACAA	AGCC	7	20
30	AAGAAAGTGA	CCATCACT	GT CCTG	CCGTC	TTTG	TCTT	GT C	TCAG	TTTC	C CT	ACAA	.CTGC	7	80
	ATTTTGTTGG	TGCAGACC	AT TGAC	CCTAT	GCCA	TGTT	CA T	CTCC	AACT	G TG	CCGT	TTCC	. 8	40

	ACCAACAT'	rg A	CATC'	TGCT'	T CC	AGGT	CACC	CAG	ACCA'	TCG	CCTT	CTTC	CA, C	AGTT	GCCT	G	900
	AACCCTGT	rc T	CTAT	GTTT'	T TG	TGGG'	TGAG	AGA'	rtcc	GCC (GGGA	TCTC	GT G	AAAA	CCCT	G	960
	AAGAACTT	GG G1	TTGC	ATCA(G CC	nggc(CCAG	TGG	GTTT(CAT	TTAC	AAGG	AG A	GAGG	GAAG	C 1	020
	TTGAAGCT	ST CO	STCT	ATGT:	r gc	TGGA(GACA	ACC'	rcag(GAG (CACT	CTCC	CT. C	TGA		1	074
5	(177) INI	FORM	ATIO	n FOI	R SE	Q ID	NO:	176:	•								
10		(B) (C) (D)	LEI TYI STI TOI	NGTH PE: 6 RANDI POLO	: 35° amin EDNE: GY: 1	7 am: o ac: SS: not:	ino a	acid: vant		•		.	•	•			
		i) MC				-						•					
		i.) SI															
	Met 1	Ala	Asp	Asp	Tyr 5	Gly	Ser	Glu	Ser	Thr 10	Ser	Ser	Met	Glu	Asp 15	Tyr	
15	Val	Asn	Phe	Asn 20	Phe	Thr	Asp	Phe	Tyr 25	Cys	Glu	, Lys	Asn	Asn 30	Val	Arg	
	Gln	Phe		Ser		Phe	Leu	Pro 40		Leu			Leu 45			Ile	
20	Val	Gly 50	Ala	Leu	Gly	Asn	Ser 55	Leu	Val	Ile	Leu	Val 60	Tyr	Trp	Tyr	Cys	
	Thr 65	Arg	Val	Lys	Thr	Met 70	Thr								Ala		
	Ala	Asp	Leu	Leu	Phe 85	Leu	Val	Thr	Leu	Pro 90		Trp	Ala	Ile	Ala 95	Ala	,
25	Ala	Asp	Gln	Trp 100	Lys	Phe	Gln	Thr	Phe 105	Met	Cys	Lys	Val	Val 110	Asn	Ser	
	Met	Tyr	Lys 115	Met	Asn	Phe	Tyr	Ser 120	Cys	Val	Leu	Leu	Ile 125	Met	Cys	Ile	
30	Ser	Val 130		Arg	Tyr	Ile	Ala 135		Ala	Gln	Ala	Met 140		Ala	His	Thr	
	Trp 145	Arg	Glu	Lys	Arg	Leu 150	Leu	Tyr	Ser	Lys	Met 155	Val	. Cys	Phe	Thr	Ile 160	
	Trp	Val	Leu	Ala	Ala 165	Ala	Leu	Cys	Ile	Pro 170	Glu	Ile	Leu	Tyr	Ser 175	Gln	
35	Ile	Lys	Glu	Glu 180	Ser	Gly	Ile	Ala	Ile	Cys	Thr	Met	Val	Tyr	Pro	Ser	

		Asp	Glu	Ser 195	Thr	Lys	Leu	Lys	Ser 200	Ala	Val	Leu	Thr	Leu 205	Lys	Val	Ile	
		Leu	Gly 210	Phe	Phe	Leu	Pro	Phe 215	Val	Val	Met	Ala	Cys 220	Cys	Tyr	Thr	Ile	
5		Ile 225	Ile	His	Thr	Leu	Ile 230	Gln	Ala	Lys	Lys	Ser 235	Ser	Lys	His	Lys	Ala 240	
		Lys	Lys	Val	Thr	Ile 245	Thr	Val	Leu	Thr	Val 250	Phe	Val	Leu	Ser	Gln 255	Phe	
10		Pro	Tyr	Asn	Cys 260	Ile	Leu	Leu	Val	Gln 265	Thr	Ile	Asp	Ala	Tyr 270	Ala	Met	
		Phe	Ile	Ser 275	Asn	Cys	Ala	Val	Ser 280	Thr	Asn	Ile	Asp	Ile 285	Cys	Phe	Gln	
		Val	Thr 290	Gln	Thr	Ile	Ala	Phe 295	Phe	His	Ser	Cys	Leu 300	Asn	Pro	Val	Leu	
15		Tyr 305	Val	Phe	Val	Gly	Glu 310	Arg	Phe	Arg	Arg	Asp 315	Leu	Val	Lys	Thr	Leu 320	
		Lys	Asn	Leu	Gly	Cys 325	Ile	Ser	Gln	Ala	Gln 330	Trp	Val	Ser	Phe	Thr 335	Arg	
20		Arg	Glu	Gly	Ser 340	Leu	Lys	Leu	Ser	Ser 345	Met	Leu	Leu	Glu	Thr 350	Thr	Ser	
		Gly	Ala	Leu 355	Ser	Leu												
	(178)) INI	FORM	OITA	v FOF	R SEÇ	O ID	NO:1	177:									
25		(i)	(B) (C)	LEI TYI STI	CE CH NGTH: PE: 1 RANDE	: 111 nucle	io ba eic a SS: s	ase p acid singl	oairs	3								
		(ii	i) Mo	DLEC	JLE T	TYPE:	: DNA	4 (ge	enomi	la)								
30		(x:	i) SI	EQUE	NCE I	DESCF	RIPTI	ON:	SEQ	ID N	10:17	77:						
	ATGG	CCTC	AT CO	SACC!	ACTCO	GGG	ccc	CAGG	GTTI	CTG	CT T	TTTAT	TCTG	G GC	TGCC	GCC	;	60
	GCGG:	rcaca	AA CI	rccc	GCCA,	A CCF	GAGO	CGCA	GAGO	CCTC	cgg (cgggc	CAACG	G GI	CGG1	GGC1	. 1	.20
	GGCG	CGGA	CG CI	CCA	GCCG1	CAC	CGCCC	CTTC	CAGA	AGCCI	rgc A	AGCT	GTGC	A TO	CAGCI	GAAG	; 1	.80
	GGGC:	rgat(CG TO	CTG	CTCTA	CAC	CGT	CGTG	GTGG	STCGI	rgg (GCT	GTGG	G CA	ACTO	CCT	5 2	24(
35	CTGG	rgcto	GG TO	ATC	GCGCC	GG1	rgcco	CGG	CTG	CACAA	ACG I	GAC	AACI	T CC	TCAT	CGGC	2 3	300

	AACCTGGCCT TGTCCGACGT GCTCATGTGC ACCGCCTGCG TGCCGCTCAC GCTGGCCTAT	360
	GCCTTCGAGC CACGCGGCTG GGTGTTCGGC GGCGGCCTGT GCCACCTGGT CTTCTTCCTG	420
	CAGCCGGTCA CCGTCTATGT GTCGGTGTTC ACGCTCACCA CCATCGCAGT GGACCGCTAC	480
	GTCGTGCTGG TGCACCCGCT GAGGCGCGCA TCTCGCTGCG CCTCAGCCTA CGCTGTGCTG	540
5	GCCATCTGGG CGCTGTCCGC GGTGCTGGCG CTGCCGCCCG CCGTGCACAC CTATCACGTG	600
	GAGCTCAAGC CGCACGACGT GCGCCTCTGC GAGGAGTTCT GGGGCTCCCA GGAGCGCCAG	660
	CGCCAGCTCT ACGCCTGGGG GCTGCTGCTG GTCACCTACC TGCTCCCTCT GCTGGTCATC	720
	CTCCTGTCTT ACGTCCGGGT GTCAGTGAAG CTCCGCAACC GCGTGGTGCC GGGCTGCGTG	780
	ACCCAGAGCC AGGCCGACTG GGACCGCGCT CGGCGCCGGC GCACCAAATG CTTGCTGGTG	840
10	GTGGTCGTGG TGGTGTTCGC CGTCTGCTGG CTGCCGCTGC ACGTCTTCAA CCTGCTGCGG	900
	GACCTCGACC CCCACGCCAT CGACCCTTAC GCCTTTGGGC TGGTGCAGCT GCTCTGCCAC	960
	TGGCTCGCCA TGAGTTCGGC CTGCTACAAC CCCTTCATCT ACGCCTGGCT GCACGACAGC 1	020
	TTCCGCGAGG AGCTGCGCAA ACTGTTGGTC GCTTGGCCCC GCAAGATAGC CCCCCATGGC 1	080
	CAGAATATGA CCGTCAGCGT GGTCATCTGA	110
15	(179) INFORMATION FOR SEQ ID NO:178:	•
20	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 369 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: (D) TOPOLOGY: not relevant 	
	(ii) MOLECULE TYPE: protein	* • •
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:178:	
	Met Ala Ser Ser Thr Thr Arg Gly Pro Arg Val Ser Asp Leu Phe Ser 1 5 10 15	•
25	Gly Leu Pro Pro Ala Val Thr Thr Pro Ala Asn Gln Ser Ala Glu Ala 20 25 30	L
	Ser Ala Gly Asn Gly Ser Val Ala Gly Ala Asp Ala Pro Ala Val Thr 35 40 45	•
30	Pro Phe Gln Ser Leu Gln Leu Val His Gln Leu Lys Gly Leu Ile Val 50 55 60	
	Leu Leu Tyr Ser Val Val Val Val Gly Leu Val Gly Asn Cys Leu 65 70 75 80	ı

	Leu	Val-	Leu	Val	Ile 85	Ala	Arg	Val	Pro	Arg 90	Leu	His	Asn	Val	Thr 95	Asn
	Phe	Leu	Ile	Gly 100	Asn	Leu	Ala	Leu	Ser 105	Asp	Val	Leu	Met	Cys 110		Ala
5	Cys	Val	Pro 115	Leu	Thr	Leu	Ala	Tyr 120	Ala	Phe	Glu	Pro	Arg 125	Gly	Trp	Val
	Phe	Gly 130	Gly	Gly	Leu	Cys	His 135	Leu	Val	Phe	Phe	Leu 140	Gln	Pro	Val	Thr
10	Val 145	Tyr	Val	Ser	Val	Phe 150	Thr	Leu	Thr	Thr	Ile 155	Ala	Val	Asp	Arg	Tyr 160
	Val	Val	Leu	Val	His 165	Pro	Leu	Arg	Arg	Ala 170	Ser	Arg	Cys	Ala	Ser 175	Ala
	Tyr	Ala	Val	Leu 180	Ala	Ile	Trp	Ala	Leu 185	Ser	Ala	Val	Leu	Ala 190	Leu	Pro
15	Pro	Ala	Val 195	His	Thr	Tyr	His	Val 200	Glu	Leu	Lys	Pro	His 205	Asp	Val	Arg
	Leu	Cys 210	Glu	Glu	Phe	Trp	Gly 215	Ser	Gln	Glu	Arg	Gln 220	Arg	Gln	Leu	Tyr
20	Ala 225	Trp	Gly	Leu	Leu	Leu 230	Val	Thr	Tyr	Leu	Leu 235	Pro	Leu	Leu	Val	Ile 240
	Leu	Leu	Ser	Tyr	Val 245	Arg	Val	Ser	Val	Lys 250	Leu	Arg	Asn	Arg	Val 255	Val
	Pro	Gly	Cys	Val 260	Thr	Gln	Ser	Gln	Ala 265	Asp	Trp	Asp	Arg	Ala 270	Arg	Arg
25			275	Lys				280					285			
	Cys	Trp 290	Leu	Pro	Leu	His	Val 295	Phe	Asn	Leu	Leu	Arg 300	Asp	Leu	Asp	Pro
30	His 305	Ala	Ile	Asp	Pro	Tyr 310	Ala	Phe	Gly	Leu	Val 315	Gln	Leu	Leu	Cys	His 320
	Trp	Leu	Ala	Met	Ser 325	Ser	Ala	Cys	Tyr	Asn 330	Pro	Phe	Ile	Tyr	Ala 335	Trp
	Leu	His	Asp	Ser 340	Phe	Arg	Glu	Glu	Leu 345	Arg	Lys	Leu	Leu	Val 350	Ala	Trp
35	Pro	Arg	Lys 355	Ile	Ala	Pro		Gly 360	Gln	Asn	Met	Thr	Val 365	Ser	Val	Val
	Ile															

(180) INFORMAT	ION FOR	SEQ II	NO:179:
----------------	---------	--------	---------

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1083 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:179:

10	ATGGACCCAG	AAGAAACTTC	AGTTTATTTG	GATTATTACT	ATGCTACGAG	CCCAAACTCT	60
	GACATCAGGG	AGACCCACTC	CCATGTTCCT	TACACCTCTG	TCTTCCTTCC	AGTCTTTTAC	120
	ACAGCTGTGT	TCCTGACTGG	AGTGCTGGGG	AACCTTGTTC	TCATGGGAGC	GTTGCATTTC	180
	AAACCCGGCA	GCCGAAGACT	GATCGACATC	TTTATCATCA	ATCTGGCTGC	CTCTGACTTC	240
	ATTTTTCTTG	TCACATTGCC	TCTCTGGGTG	GATAAAGAAG	CATCTCTAGG	ACTGTGGAGG	3.00
15	ACGGGCTCCT	TCCTGTGCAA	AGGGAGCTCC	TACATGATCT	CCGTCAATAT	GCACTGCAGT	360
	GTCCTCCTGC	TCACTTGCAT	GAGTGTTGAC	CGCTACCTGG	CCATTGTGTG	GCCAGTCGTA	420
	TCCAGGAAAT	TCAGAAGGAC	AGACTGTGCA	TATGTAGTCT	GTGCCAGCAT	CTGGTTTATC	480
	TCCTGCCTGC	TGGGGTTGCC	TACTCTTCTG	TCCAGGGAGC	TCACGCTGAT	TGATGATAAG	540
	CCATACTGTG	CAGAGAAAAA	GGCAACTCCA	ATTAAACTCA	TATGGTCCCT	GGTGGCCTTA	600
20	ATTTTCACCT	TTTTTGTCCC	TTTGTTGAGC	ATTGTGACCT	GCTACTGTTG	CATTGCAAGG	660
	AAGCTGTGTG	CCCATTACCA	GCAATCAGGA	AAGCACAACA	AAAAGCTGAA	GAAATCTAAG	720
	AAGATCATCT	TTATTGTCGT	GGCAGCCTTT	CTTGTCTCCT	GGCTGCCCTT	CAATACTTTC	780
·	AAGTŢCCTGG	CCATTGTCTC	TGGGTTGCGG	CAAGAACACT	ATTTACCCTC	AGCTATTCTT	840
÷	CAGCTTGGTA	TGGAGGTGAG	TGGACCCTTG	GCATTTGCCA	ACAGCTGTGT	CAACCCTTTC	900
25	ATTTACTATA	TCTTCGACAG	CTACATCCGC	CGGGCCATTG	TCCACTGCTT	GTGCCCTTGC	960
		ATGACTTTGG					1020
		TCATTCATGC	AGAAGATTTT	GCCAGGAGGA	GGAAGAGGTC	TGTGTCACTC	1080
	TAA						1083

(181) INFORMATION FOR SEQ ID NO:180:

(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 360 amino acids

Commence of the south for an

		(C)	ST	PE: & RANDI POLOC	EDNES	SS:		vant								
	(i:	L) MC	DLECT	JLE T	TYPE:	pro	oteir	ı								
5	(xi	i) SI	EQUE	ICE I	DESCI	RIPTI	ON:	SEQ	ID 1	NO:18	30:					
	Met 1	Asp	Pro	Glu	Glu 5	Thr	Ser	Val	Tyr	Leu 10	Asp	Tyr	Tyr	Tyr	Ala 15	Thr
	Ser	Pro	Asn	Ser 20	Asp	Ile	Arg	Glu	Thr 25	His	Ser	His	Val	Pro 30	Tyr	Thr
10	Ser	Val	Phe 35	Leu	Pro	Val	Phe	Tyr 40	Thr	Ala	Val	Phe	Leu 45	Thr	Gly	Val
	Leu	Gly 50	Asn	Leu	Val	Leu	Met 55	Gly	Ala	Leu	His	Phe 60	Lys	Pro	Gly	Ser
15	Arg 65	Arg	Leu	Ile	Asp	Ile 70	Phe	Ile	Ile	Asn	Leu 75	Ala	Ala	Ser	Asp	Phe 80
	Ile	Phe	Leu	Val	Thr 85	Leu	Pro	Leu	Trp	Val 90	Asp	Lys	Glu	Ala	Ser 95	Leu
	Gly	Leu	Trp	Arg 100	Thr	Gly	Ser	Phe	Leu 105	Cys	Lys	Gly	Ser	Ser 110	Tyr	Met
20	Ile	Ser	Val 115	Asn	Met	His	Cys	Ser 120	Val	Leu	Leu	Leu	Thr 125	Cys	Met	Ser
	Val	Asp 130	Arg	Tyr	Leu	Ala	Ile 135	Val	Trp	Pro	Val	Val 140	Ser	Arg	Lys	Phe
25	Arg 145	Arg	Thr	Asp	Cys	Ala 150	Tyr	Val	Val	-Cys	Ala 155	Ser	Ile	Trp	Phe	Ile 160
	Ser	Cys	Leu	Leu	Gly 165	Leu	Pro	Thr	Leu	Leu 170	Ser	Arg	Glu	Leu	Thr 175	Leu
	Ile	Asp	Asp	Lys 180	Pro	Tyr	Cys	Ala	Glu 185	Lys	Lys	Ala	Thr	Pro 190	Ile	Lys
30	Leu	Ile	Trp 195	Ser	Leu	Val	Ala	Leu 200	Ile	Phe	Thr	Phe	Phe 205	Val	Pro	Leu
	Leu	Ser 210	Ile	Val	Thr	Cys	Tyr 215	Cys	Cys	Ile	Ala	Arg 220	Lys	Leu	Cys	Ala
35	His 225	Tyr	Gln	Gln	Ser	Gly 230	Lys	His	Asn	Lys	Lys 235	Leu	Lys	Lys	Ser	Lys 240
	Lys	Ile	Ile	Phe	Ile 245	Val	Val	Ala	Ala	Phe 250	Leu	Val	Ser	Trp	Leu 255	Pro

	Phe i	Asn Thr	Phe 260	Lys	Phe	Leu	Ala	Ile 265	Val	Ser	Gly	Leu	Arg 270	Gln	Glu	
	His '	Tyr Leu 275		Ser	Ala	Ile	Leu 280	Gln	Leu	Gly	Met	Glu 285	Val	Ser	Gly	
5		Leu Ala 290	Phe .	Ala	Asn	Ser 295	Cys	Val	Asn	Pro	Phe 300	Ile	Tyr	Tyr	Ile	
	Phe 2	Asp Ser	Tyr		Arg 310	Arg	Ala	Ile	Val	His 315	Cys	Leu	Cys	Pro	Cys 320	
10	Leu :	Lys Asn	-	Asp 325	Phe	Gly	Ser	Ser	Thr 330	Glu	Thr	Ser	Asp	Ser 335	His	
	Leu '	Thr Lys	Ala :	Leu	Ser	Thr	Phe	Ile 345	His	Ala	Glu	Asp	Phe 350	Ala	Arg	
	Arg :	Arg Lys 355	Arg	Ser	Val	Ser	Leu 360									
15	(182) INF	ORMATIO	N FOR	SEQ	ID	мо:1	.81:									
20	(ii	SEQUEN (A) LE (B) TY (C) ST (D) TO) MOLEC) SEQUE	NGTH: PE: n' RANDE POLOG' ULE T'	102 ucle DNES Y: l YPE:	0 baic assinea	ise pacid singl ar	enomi	.c)	JO:18	31:						
	ATGAATGGC	C TTGAA	GTGGC	TCC	CCCA	AGGT	CTGF	TCAC	CA A	ACTTO	TCCC	T GO	CCAC	CGGC	.	6
	GAGCAATGT	G GCCAG	GAGAC	GCC	ACTO	GAG	AACA	TGCT	GT I	rcgcc	TCCI	T CI	TACCI	TCT	3 1	12
25	GATTTTATC	C TGGCT	TTAGT	TGG	CAAT	CACC	CTG	CTCI	GT G	GCTI	TTC	T C	GAGA	ACCAC	: 1	18
	AAGTCCGGG	A CCCCG	GCCAA	CGT	GTTC	CTG	ATGO	CATCI	rgg (CCGT	GCCG	SA CI	TGT	GTGC	: 2	24
	GTGCTGGTC	C TGCCC	ACCCG	CCT	GGTC	CTAC	CACI	TCTC	CTG G	GAAC	CACI	G GC	CATI	TGGG	3	30
	GAAATCGCA'	T GCCGT	CTCAC	CGG	CTTC	CTC	TTCI	CACCI	CA A	CATO	TAC	C C	AGCAI	CTAC	2 3	36
	TTCCTCACC	T GCATC	AGCGC	CGA	.CCG1	TTC	CTGG	CCAT	TTG 1	rgcac	CCGG	T C	AGTO	CCTC	2 4	12
30	AAGCTCCGC	A GGCCC	CTCTA	CGC	ACAC	CTG	GCC1	GTGC	CT 1	CCT	TGGG	T GO	TGGT	rggci	. 4	18
	GTGGCCATG	g ccccg	CTGCT	GGT	GAGO	CCA	CAGA	ACCGI	rgc <i>I</i>	GAC	AACC	A C	CGGI	rggro	: :	54
	TGCCTGCAG	C TGTAC	CGGGA	GAA	.GGCC	TCC	CACC	CATGO	ccc 1	GGTG	STCCC	T GO	CAGI	rggco	: 6	50
	TTCACCTTC	C CGTTC	ATCAC	CAC	GGTC	ACC	TGCT	יארריו	יפר יו	CATC	אירכ	יה כז	יפירט	racaa		561

133

CAGGGCCTGC	GTGTGGAGAA	GCGCCTCAAG	ACCAAGGCAA	AACGCATGAT	CGCCATAGTG	720
CTGGCCATCT	TCCTGGTCTG	CTTCGTGCCC	TACCACGTCA	ACCGCTCCGT	CTACGTGCTG	780
CACTACCGCA	GCCATGGGGC	CTCCTGCGCC	ACCCAGCGCA	TCCTGGCCCT	GGCAAACCGC	840
ATCACCTCCT	GCCTCACCAG	CCTCAACGGG	GCACTCGACC	CCATCATGTA	TTTCTTCGTG	900
GCTGAGAAGT	TCCGCCACGC	CCTGTGCAAC	TTGCTCTGTG	GCAAAAGGCT	CAAGGGCCCG	960
CCCCCAGCT	TCGAAGGGAA	AACCAACGAG	AGCTCGCTGA	GTGCCAAGTC	AGAGCTGTGA	1020
(183) INFOR	MATION FOR	SEQ ID NO:1	L82:		•	

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 339 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:

5

- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:182:
- Val Leu Val Leu Pro Thr Arg Leu Val Tyr His Phe Ser Gly Asn His 85 90 95
 - Trp Pro Phe Gly Glu Ile Ala Cys Arg Leu Thr Gly Phe Leu Phe Tyr 100 105 110
- Leu Asn Met Tyr Ala Ser Ile Tyr Phe Leu Thr Cys Ile Ser Ala Asp
 115 120 125
 - Arg Phe Leu Ala Ile Val His Pro Val Lys Ser Leu Lys Leu Arg Arg 130 135 140
 - Pro Leu Tyr Ala His Leu Ala Cys Ala Phe Leu Trp Val Val Val Ala 145 150 155 160
- Val Ala Met Ala Pro Leu Leu Val Ser Pro Gln Thr Val Gln Thr Asn

			•		165					170	::: .		v -		175	
	His	Thr	Val	Val 180	Суз	Leu	Gln	Leu	Tyr 185	Arg	Glu	Lys	Ala	Ser 190	His	His
5	Ala	Leu	Val 195	Ser	Leu	Ala	Val	Ala 200	Phe	Thr		Pro	Phe 205	Ile	Thr	Thr
	Val	Thr 210	Cys	Tyr	Leu	Leu	Ile 215	Ile	Arg	Ser	Leu	Arg 220	Gln	Gly	Leu	Arg
	Val 225	Glu	Lys	Arg	Leu	Lys 230	Thr	Lys	Ala	Lys	Arg 235	Met	Ile	Ala	Ile	Val 240
10	Lev	Ala	Ile	Phe	Leu 245	Val	Cys	Phe	Val	Pro 250	Tyr	His	Val	Asn	Arg 255	Ser
	Val	Tyr		Leu 260	His	Tyr	Arg	Ser	His 265	Gly	Ala	Ser	Cys	Ala 270	Thr	Gln
15	Arg	Ile	Leu 275	Ala	Leu	Ala	Asn	Arg 280	Ile	Thr	Ser	Cys	Leu 285	Thr	Ser	Leu
	Asn	Gly 290	Ala	Leu	Asp	Pro	Ile 295	Met	Tyr	Phe	Phe	Val 300	Ala	Glu	Lys	Phe
	Arg 305	His	Ala	Leu	Cys	Asn 310	Leu	Leu	Cys	Gly	Lys 315	Arg	Leu	Lys		Pro 320
20	Pro	Pro	Ser	Phe	Glu 325	Gly	Lys	Thr	Asn	Glu 330	Ser	Ser	Leu	Ser	Ala 335	Lys
	Ser	Glu	Leu													
	(183) IN	FORMA	TION	FOR	SEQ	D	NO:1	.83 :								
25	(i	(B)	UENC LEN TYP STR TOP	GTH: E: n	996 ucle	bas ic a S: s	e pa cid ingl	irs						. •		
30	(i:	i) MC	LECU	LE T	YPE:	DNA	. (ge	nomi	c)							
	(x:	i) SE	QUEN	CE D	ESCR	IPTI	ON:	SEQ	ID N	0:18	3:					
	ATGATCAC	CC TG	AACA	ATCA	AGA	TCAA	CCT	GTCC	CTTT	TA A	CAGC	TCAC	A TC	CAGA	TGAA	60
	TACAAAAT	rg ca	GCCC	TTGT	CTT	CTAT.	AGC	TGTA	TCTT	CA T	AATT	GGAT	T AT	TTGT	TAAC	120
	ATCACTGC	AT TA	TGGG	TTTT	CAG	TTGT.	ACC .	ACCA	AGAA	GA G	AACC.	ACGG	T AA	CCAT	CTAT	180
35	ATGATGAA	rg Tg	GCAT	TAGT	GGA	CTTG.	ATA	TTTA	TAAT	GA C	TTTA	CCCT	T TC	GAAT	GTTT	240

	TATTATGC	A AA	AGAT	GAATO	G GC	CATT	TGGA	GAG	TACT	TCT	GCCA	GATT	CT T	GGAG	CTCT	C	300
	ACAGTGTT	TT A	CCCA	AGCAT	TG	CTTT	ATGG	CTT	CTTG	CCT	TATT	TAGT	GC T	GACA	GATA	С	360
	ATGGCCAT	rg T	ACAG	CCGAZ	A GT	ACGC	CAAA	GAA	CTTA	AAA	ACAC	GTGC.	AA A	GCCG	TGCT	G	420
	GCGTGTGT	G G	AGTC'	TGGAI	'AA'	TGAC	CCTG	ACC	ACGA	CCA	CCCC'	rctg	CT A	CTGC'	TCTA:	r	480
5	AAAGACCC	AG A'	AAAT	GACTO	CA	CTCC	CGCC	ACC'	rgcc'	rca .	AGAT'	TTCT	GA C	ATCA'	TCTA	r	540
	CTAAAAGCT	rg To	GAAC	GTGCT	' GA	ACCT	CACT	CGA	CTGA	CAT	TTTT	TTTC'	rt G	ATTC	CTTT	3	600
	TTCATCATC	SA T'	rggg:	TGCT	CT	TGGT	CATT	ATT	CATA	ATC	TCCT	rcac(GG C	AGGA	CGTC	r	660
	AAGCTGAA <i>I</i>	AC C	CAAA	GTCAA	GG	AGAA	GTCC	AAA	AGGA:	rca	TCAT	CACG	CT G	CTGG'	TGCA	3	720
	GTGCTCGTC	CT G	CTTT	ATGCC	CT	TCCA(CATC	TGT	TTCG	CTT	TCCT	GATG	CT G	GGAA(CGGGG	3	780
10	GAGAATAGT	A TI	CAAT	CCCTG	GG	GAGC	CTTT	ACC	ACCT	rcc	TCAT	GAAC	CT C	AGCA	CGTG	Г	840
	CTGGATGTC	SA T	rctc:	TACTA	CA!	rcgt'	TTCA	AAA	CAAT	TTC .	AGGC:	rcga(T C	ATTA	GTGT	3	900
	ATGCTATAC	CC G	TAAT	racci	TC	GAAG	CATG	CGC	AGAA	AAA	GTTT	CCGA:	rc To	GTA(GTCT#	A	960
	AGGTCACTA	AA G	CAAT	AAATA	CAC	GTGA	AATG	TTA:	rga								996
	(185) INF	FORM	OITA	N FOR	SE	Q ID	NO:	184:									
15	(i)	(A) (B) (C)	LEI TYI	CE CH NGTH: PE: a RANDE	33: mino DNES	lam: cac: SS:	ino a id	acid									
20	(ii	.) MC	OLECT	JLE I	YPE	: pro	oteir	1									
	(xi	.) SI	EQUE	NCE D	ESCI	RIPT	ION:	SEQ	ID 1	10:1	84:			•			
	Met 1	Ile	Thr	Leu	Asn 5	Asn	Gln	Asp	Gln	Pro 10	Val	Pro	Phe	Asn	Ser 15	Ser	
25	His	Pro	Asp	Glu 20	Tyr	Lys	Ile	Ala	Ala 25	Leu	Val	Phe	Tyr	Ser 30	Cys	Ile	
	Phe	Ile	Ile 35	Gly	Leu	Phe	Val	Asn 40	Ile	Thr	Ala	Leu	Trp 45	Val	Phe	Ser	
	Cys	Thr 50	Thr	Lys	Lys	Arg	Thr 55	Thr	Val	Thr	Ile	Tyr 60	Met	Met	Asn	Val	
30	Ala 65	Leu	Val	Asp	Leu	Ile 70	Phe	Ile	Met	Thr	Leu 75	Pro	Phe	Arg	Met	Phe 80	
	Tyr	Tyr	Ala	Lys	Asp 85	Glu	Trp	Pro	Phe	Gly 90	Glu	Tyr	Phe	Cys	Gln 95	Ile	

		Leu	Gly	Ala	Leu 100	Thr	Val	Phe	Tyr	Pro 105	Ser	Ile	Ala	Leu	Trp 110	Leu	Leu
	•	Ala	Phe	Ile 115	Ser	Ala	Asp	Arg	Tyr 120	Met	Ala	Ile	Val	Gln 125	Pro	Lys	Tyr
5		Ala	Lys 130	Glu	Leu	Lys	Asn	Thr 135	Cys	Lys	Ala	Val	Leu 140	Ala	Cys	Val	Gly
	•	Val 145	Trp	Ile	Met	Thr	Leu 150	Thr	Thr	Thr	Thr	Pro 155		Leu	Leu	Leu	Tyr 160
10		Lys	Asp	Pro	Asp	Lys 165	Asp	Ser	Thr	Pro	Ala 170	Thr	Cys	Leu	Lys	Ile 175	Ser
		Asp	Ile	Ile	Tyr 180	Leu	Lys	Ala	Val	Asn 185	Val	Leu	Asn	Leu	Thr 190	Arg	Leu
	• • .	Thr	Phe	Phe 195	Phe	Leu	Ile	Pro	Leu 200	Phe	Ile	Met	Ile	Gly 205	CAè	Tyr	Leu
15		Val	Ile 210	Ile	His	Asn	Leu	Leu 215	His	Gly	Arg	Thr	Ser 220	Lys	Leu	Lys	Pro
	÷	Lys 225	Val	Lys	Glu	Lys	Ser 230	Lys	Arg	Ile	Ile	Ile 235	Thr	Leu	Leu	Val	Gln 240
20		Val	Leu	Val	Cys	Phe 245	Met	Pro	Phe	His	Ile 250	Cys	Phe	Ala	Phe	Leu 255	Met
		Leu	Gly	Thr	Gly 260	Glu	Asn	Ser	Tyr	Asn 265	Pro	Trp	Gly	Ala	Phe 270	Thr	Thr
		Phe	Leu	Met 275	Asn	Leu	Ser	Thr	Cys 280		Asp	Val	Ile	Leu 285	Tyr	Tyr	Ile
25		Val	Ser 290	Lys	Gln	Phe	Gln	Ala 295	Arg	Val	Ile	Ser	Val 300	Met	Leu	Tyr	Arg
		Asn 305	Tyr	Leu	Arg	Ser	Met 310	Arg	Arg	Lys	Ser	Phe 315	Arg	Ser	Gly	Ser	Leu 320
30		Arg	Ser	Leu	Ser	Asn 325	Ile	Asn	Ser	Glu	330	Leu					
	(186)) IN:	FORM	ATIO	N FO	R SE	αI Ç	NO:	185:		**		,				
		(i) SE	-					CS: pair:	•				,	•		
35			(B (C) TY:) ST:) TO:	PE: :	nucle EDNE:	eic a SS: a	acid sing		3							

(ii) MOLECULE TYPE: DNA (genomic)

137

	(xi)	SEQUENCE DI	ESCRIPTION:	SEQ ID NO:	185:	· •	
	ATGCCCTCTG	TGTCTCCAGC	GGGCCCTCG	GCCGGGGCAG	TCCCCAATGC	CACCGCAGTG	60
	ACAACAGTGC	GGACCAATGC	CAGCGGGCTG	GAGGTGCCCC	TGTTCCACCT	GTTTGCCCGG	120
	CTGGACGAGG	AGCTGCATGG	CACCTTCCCA	GGCCTGTGCG	TGGCGCTGAT	GGCGGTGCAC	180
5	GGAGCCATCT	TCCTGGCAGG	GCTGGTGCTC	AACGGGCTGG	CGCTGTACGT	CTTCTGCTGC	240
	CGCACCCGGG	CCAAGACACC	CTCAGTCATC	TACACCATCA	ACCTGGTGGT	GACCGATCTA	300
	CTGGTAGGGC	TGTCCCTGCC	CACGCGCTTC	GCTGTGTACT	ACGGCGCCAG	GGGCTGCCTG	360
	CGCTGTGCCT	TCCCGCACGT	CCTCGGTTAC	TTCCTCAACA	TGCACTGCTC	CATCCTCTTC	420
	CTCACCTGCA	TCTGCGTGGA	CCGCTACCTG	GCCATCGTGC	GGCCCGAAGG	CTCCCGCCGC	480
10	TGCCGCCAGC	CTGCCTGTGC	CAGGGCCGTG	TGCGCCTTCG	TGTGGCTGGC	CGCCGGTGCC	540
	GTCACCCTGT	CGGTGCTGGG	CGTGACAGGC	AGCCGGCCCT	GCTGCCGTGT	CTTTGCGCTG	60,0
	ACTGTCCTGG	AGTTCCTGCT	GCCCCTGCTG	GTCATCAGCG	TGTTTACCGG	CCGCATCATG	660
	TGTGCACTGT	CGCGGCCGGG	TCTGCTCCAC	CAGGGTCGCC	AGCGCCGCGT	GCGGGCCAAG	720
	CAGCTCCTGC	TCACGGTGCT	CATCATCTTT	CTCGTCTGCT	TCACGCCCTT	CCACGCCCGC	780
15	CAAGTGGCCG	TGGCGCTGTG	GCCCGACATG	CCACACCACA	CGAGCCTCGT	GGTCTACCAC	840
	GTGGCCGTGA	CCCTCAGCAG	CCTCAACAGC	TGCATGGACC	CCATCGTCTA	CTGCTTCGTC	900
	ACCAGTGGCT	TCCAGGCCAC	CGTCCGAGGC	CTCTTCGGCC	AGCACGGAGA	GCGTGAGCCC	960
	AGCAGCGGTG	ACGTGGTCAG	CATGCACAGG	AGCTCCAAGG	GCTCAGGCCG	TCATCACATC	1020
	CTCAGTGCCG	GCCCTCACGC	CCTCACCCAG	GCCCTGGCTA	ATGGGCCCGA	GGCTTAG	1077
20	(187) INFO	RMATION FOR	SEQ ID NO:	186:			

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 358 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
- - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:186:

Met Pro Ser Val Ser Pro Ala Gly Pro Ser Ala Gly Ala Val Pro Asn 1 5 10 15

Ala Thr Ala Val Thr Thr Val Arg Thr Asn Ala Ser Gly Leu Glu Val 20 25 30

	Pr	o Le	ı Phe 35	His	Leu	Phe	Ala	Arg 40	Leu	Asp	Glu	Glu	Leu 45	His	Gly	Thr
	Ph	e Pro	o Gly	Leu	Cys	Val	Ala 55	Leu	Met	Ala	Val	His 60	Gly	Ala	Ile	Phe
5	Le 65	u Ala	a Gly	Leu	Val	Leu 70	Asn	Gly	Leu	Ala	Leu 75	Tyr	Val	Phe	Cys	Cys 80
	Ar	g Thi	Arg	Ala	Lys 85	Thr	Pro	Ser	Val	Ile 90	Tyr	Thr	Ile	Asn	Leu 95	Val
10	Va	l Thi	Asp	Leu 100	Leu	Val	Gly	Leu	Ser 105		Pro	Thr	Arg	Phe 110	Ala	Val
	Ту	т Туі	: Gly 115		Arg	Gly	Cys	Leu 120	Arg	Cys	Ala	Phe	Pro 125	His	Val	Leu
	Gl	у Туз 130		Leu	Asn	Met	His 135	Cys	Ser	Ile	Leu	Phe 140	Leu	Thr	Cys	Ile
15	Су 14	s Val	. Asp	Arg	Tyr	Leu 150	Ala	Ile	Val	Arg	Pro 155	Glu	Gly	Ser	Arg	Ala 160
	Су	s Arg	Gln	Pro	Ala 165	Суѕ	Ala	Arg	Ala	Val 170	Cys	Ala	Phe	Val	Trp 175	Leu
20	Al	a Ala	Gly	Ala 180	Val	Thr	Leu	Ser	Val 185	Leu	Gly	Val	Thr	Gly 190	Ser	Arg
	Pr	o Cys	Cys 195	Arg	Val	Phe	Ala	Leu 200	Thr	Val	Leu	Glu	Phe 205	Leu	Leu	Pro
		u Leu 210	١				215					220	•	•	• •	
25	22					230					235					240
		n Leu			245					250					255	
30		e His		260					265					270		
		s Thr	275					280					285			
		n Ser 290					295					300			_	
35	30					310					315					320
	Se:	r Ser	Gly	Asp	Val	Val	Ser	Met	His	Ara	Ser	Ser	Lvs	Glv	Ser	Glv

באפרייות בשת מממימה ו

	·		325	1 A 33	30 - 1	335	
<i>A</i> .	Arg Hi	is His Ile 1 340	Leu Ser Ala	Gly Pro Hi	s Ala Leu 1	Thr Gln Ala 350	Leu
5	Ala As	sn Gly Pro (Glu Ala	ar e j	•		
	(188) INFOR	MATION FOR	SEQ ID NO:	187:			
10	; ' ((A) LENGTH: (B) TYPE: no	ARACTERISTIC 1050 base p scleic acid DNESS: sing (: linear	pairs		; .	
	(ii)	MOLECULE TY	(PE: DNA (ge	enomic)			
	(xi)	SEQUENCE DE	ESCRIPTION:	SEQ ID NO:	187:	I to a state of	
	ATGAACTCCA	CCTTGGATGG	TAATCAGAGC	AGCCACCCTT	TTTGCCTCTT	GGCATTTGGC	6,0
15	TATTTGGAAA	CTGTCAATTT	TTGCCTTTTG	GAAGTATTGA	TTATTGTCTT	TCTAACTGTA	120
	TTGATTATTT	CTGGCAACAT	CATTGTGATT	TTTGTATTTC	ACTGTGCACC	TTTGTTGAAC	180
	CATCACACTA	CAAGTTATTT	TATCCAĜACT	ATGGCATATG	CTGACCTTTT	TGTTGGGGTG	240
	AGCTGCGTGG	TCCCTTCTTT	ATCACTCCTC	CATCACCCCC	TTCCAGTAGA	. GGAGTCCTTG	300
	ACTTGCCAGA	TATTTGGTTT	TGTAGTATCA	GTTCTGAAGA	GCGTCTCCAT	GCTTCTCTG	360
20	GCCTGTATCA	GCATTGATAG	ATACATTGCC	ATTACTAAAC	CTTTAACCTA	TAATACTCTG	420
	GTTACACCCT	GGAGACTACG	CCTGTGTATT	TTCCTGATTT	GGCTATACTC	GACCCTGGTC	480
	TTCCTGCCTT	CCTTTTTCCA	CTGGGGCAAA	CCTGGATATC	ATGGAGATGT	GTTTCAGTGG	540
	TGTGCGGAGT	CCTGGCACAC	CGACTCCTAC	TTCACCCTGT	TCATCGTGAT	GATGTTATAT	600
	GCCCCAGCAG	CCCTTATTGT	CTGCTTCACC	TATTTCAACA	TCTTCCGCAT	CTGCCAACAG	660
25	CACACAAAGG	ATATCAGCGA	AAGGCAAGCC	CGCTTCAGCA	GCCAGAGTGG	GGAGACTGGG	720
	GAAGTGCAGG	CCTGTCCTGA	TAAGCGCTAT	AAAATGGTCC	TGTTTCGAAT	CACTAGTGTA	780
	TTTTACATCC	TCTGGTTGCC	ATATATCATC	TACTTCTTGT	TGGAAAGCTC	CACTGGCCAC	840
	AGCAACCGCT	TCGCATCCTT	CTTGACCACC	TGGCTTGCTA	TTAGTAACAG	TTTCTGCAAC	900
	TGTGTAATTT	ATAGTCTCTC	CAACAGTGTA	TTCCAAAGAG	GACTAAAGCG	CCTCTCAGGG	960
30	GCTATGTGTA	CTTCTTGTGC	AAGTCAGACT	ACAGCCAACG	ACCCTTACAC	AGTTAGAAGC	1020
	AAAGGCCCTC	TTAATGGATG	TCATATCTGA				1050

	(189)	INF	ORMA	TION	FOR	SEÇ	D ID			5.	. •.			•••			
5		(i)	(A) (B) (C)	LEN TYI STR	IGTH: PE: a RANDE	349 minc DNES	TERI ami aci S:	.no a .d	cids	.	-	•					
		(ii	.) MC	LECU	TE 1	YPE:	pro	teir	1								
		(xi	.) SE	EQUEN	ICE I	ESCF	RIPTI	ON:	SEQ	ID N	10:18	8:					
10		Met 1	Asn	Ser	Thr	Leu 5	Asp	Gly	Asn	Gln	Ser 10	Ser	His	Pro	Phe	Cys 15	Leu
		Leu	Ala	Phe	Gly 20	Tyr	Leu	Glu	Thr	Val 25	Asn	Phe	Cys	Leu	Leu 30	Glu	Val
	, '	Leu	Ile	Ile 35	Val	Phe	Leu		Val 40	Leu	Ile	Ile	Ser	Gly 45	Asn	Ile	Ile
15		Val	Ile 50	Phe	Val	Phe	His	Суs 55	Ala	Pro	Leu	Leu	Asn 60	His	His	Thr	Thr
		Ser 65	Tyr	Phe	Ile	Gln	Thr 70	Met	Ala	Tyr	Ala	Asp 75	Leu	Phe	Val	Gly	Val 80
20		Ser	Cys	Val	Val	Pro 85	Ser	Leu	Ser	Leu	Leu 90	His	His	Pro	Leu	Pro 95	Val.
		Glu	Glu	Ser	Leu 100	Thr	Cys	Gln	Ile	Phe 105	Gly	Phe	Val	Val	Ser 110	Val	Leu
		Lys	Ser	Val	Ser	Met	Ala		120	Ala	Cys			Ile 125		Arg	Tyr
25		Ile	Ala 130		Thr	Lys	Pro		Thr	Tyr	Asn					Pro	Trp
																	Val 160
30		Phe	Leu	Pro	Ser	Phe 165	Phe	His	Trp	Gly	Lys 170	Pro	Gly	Tyr	His	Gly 175	Asp
		Val	Phe	Gln	Trp 180	Cys	Ala	Glu	Ser	Trp 185	His	Thr	Asp	Ser	Tyr 190	Phe	Thr
•		Leu	Phe	Ile 195	Val	Met	Met	Leu	Tyr 200	Ala	Pro	Ala	Ala	Leu 205	Ile	Val	Cys
35		Phe	Thr 210	-	Phe	Asn	Ile	Phe 215	Arg	Ile	Cys	Gln	Gln 220	His	Thr	Lys	Asp

	Ile 225	Ser	Glu	Arg	Gln	Ala 230	Arg	Phe	Ser	Ser	Gln 235	Ser	Gly	Glu	Thr	Gly 240	
	Glu	Val	Gln	Ala	Cys 245	Pro	Asp	Lys	Arg	Tyr 250	Lys	Met	Val	Leu	Phe 255	Arg	
5	Ile	Thr	Ser	Val 260	Phe	Tyr	Ile	Leu	Trp 265	Leu	Pro	Tyr	Ile	Ile 270	Tyr	Phe	
	Leu	Leu	Glu 275	Ser	Ser	Thr	Gly	His 280	Ser	Asn	Arg	Phe	Ala 285	Ser	Phe	Leu	
10	Thr	Thr 290	Trp	Leu	Ala	Ile	Ser 295	Asn	Ser	Phe	Cys	Asn 300	Cys	Val	Ile	Tyr	
	Ser 305	Leu	Ser	Asn	Ser	Val 310	Phe	Gln	Arg	Gly	Leu 315		Arg	Leu	Ser	Gly 320	
	Ala	Met	Cys	Thr	Ser 325	Cys	Ala	Ser	Gln	Thr 330	Thr	Ala	Asn	Asp	Pro 335	Tyr	
15	Thr	Val	Arg	Ser 340	Lys	Gly	Pro	Leu	Asn 345	Gly	Cys	His	Ile				ŧ,
	(190) IN	FORM	TION	FOR	SEÇ	ID	NO:1	89:								.*	
20	(i)	(A) (B) ·(C)	LEN TYP STR TOP	IGTH: PE: n LANDE	130 ucle DNES	2 ba ic a S: s	se p cid ingl	airs	• .					··			
	· (ii	.) MC	LECU	LE T	YPE :	DNA	(ge	nomi	c)								
	(xi) SE	QUEN	CE D	ESCR	IPTI	ON:	SEQ	ID N	0:18	9 :						
25	ATGTGTTTT	T CI	CCCA	TTCT	GGA	AATC	AAC	ATGC	AGTC	TG A	ATCT	AACA	T TA	.CAGT	GCGA		60
	GATGACATI	TA D	'GACA	TCAA	CAC	CAAT	ATG	TACC	AACC	AC T	ATCA	TATC	C GT	TAAG	CTTT	1	20
	CAAGTGTCT	C TC	ACCG	GATT	TCT	TATG	TTA	GAAA	TTGT	GT T	GGGA	CTTG	G CA	GCAA	CCTC	1	80
	ACTGTATTG	G TA	CTTT	ACTG	CAT	GAAA	TCC .	AACT	TAAT	CA A	CTCT	GTCA	G TA	ACAT	TATT	2	40
	ACAATGAAT	C TT	CATG	TACT	TGA'	TGTA.	ATA .	ATTT	GTGT	GG G	ATGT.	ATTC	C TC	TAAC	TATA	3	00
30	GTTATCCTT	C TG	CTTT	CACT	GGA	GAGT.	AAC .	ACTG	CTCT	CA T	TTGC	TGTT	r cc	ATGA	GGCT	3	60
	TGTGTATCT	T TT	GCAA	GTGT	CTC	AACA	GCA :	ATCA	ACGT	тт т	TGCT.	ATCA	C TT	TGGA	CAGA	4	20
	TATGACATO	т ст	GTAA.	AACC	TGC	AAAC	CGA 2	ATTC'	TGAC	AA T	GGGC.	AGAG	C TG	TAAT	GTTA	4	80
	ATGATATCC	A TT	TGGA'	TTTT	TTC	TTTT'	TTC '	rctt	rccr	GA T	TCCT'	TTTA'	r TG	AGGT	TAAA	5.	40
	TTTTTCAGT	C TT	CAAA	GTGG	AAA'	TACC'	TGG (GAAA	ACAA	GA C	ACTT'	TAT(G TG	TCAG'	TACA	6	00

	AATGAATACT ACACTGAACT GGGAATGTAT TATCACCTGT TAGTACAGAT CCCAATATTC	660
	TTTTTCACTG TTGTAGTAAT GTTAATCACA TACACCAAAA TACTTCAGGC TCTTAATATT	720
	CGAATAGGCA CAAGATTTTC AACAGGGCAG AAGAAGAAAG CAAGAAAGAA AAAGACAATT	780
	TCTCTAACCA CACAACATGA GGCTACAGAC ATGTCACAAA GCAGTGGTGG GAGAAATGTA	840
5	GTCTTTGGTG TAAGAACTTC AGTTTCTGTA ATAATTGCCC TCCGGCGAGC TGTGAAACGA	900
	CACCGTGAAC GACGAGAAG ACAAAAGAGA GTCAAGAGGA TGTCTTTATT GATTATTTCT	960
	ACATTTCTTC TCTGCTGGAC ACCAATTTCT GTTTTAAATA CCACCATTTT ATGTTTAGGC	1020
	CCAAGTGACC TTTTAGTAAA ATTAAGATTG TGTTTTTTAG TCATGGCTTA TGGAACAACT	1080
	ATATTTCACC CTCTATTATA TGCATTCACT AGACAAAAAT TTCAAAAGGT CTTGAAAAGT	1140
10	AAAATGAAAA AGCGAGTTGT TTCTATAGTA GAAGCTGATC CCCTGCCTAA TAATGCTGTA	1200
	ATACACAACT CTTGGATAGA TCCCAAAAGA AACAAAAAA TTACCTTTGA AGATAGTGAA	1260
	ATAAGAGAAA AACGTTTAGT GCCTCAGGTT GTCACAGACT AG	1302
	(191) INFORMATION FOR SEQ ID NO:190:	
15	(i) SEQUENCE CHARACTERISTICS:	
1)	(A) LENGTH: 433 amino acids (B) TYPE: amino acid	
	(C) STRANDEDNESS: (D) TOPOLOGY: not relevant	
	(ii) MOLECULE TYPE: protein	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:190:	
	Met Cys Phe Ser Pro Ile Leu Glu Ile Asn Met Gln Ser Glu Ser As 1 5 10 15	sn į
	Ile Thr Val Arg Asp Asp Ile Asp Asp Ile Asn Thr Asn Met Tyr Gl	ln
25	Pro Leu Ser Tyr Pro Leu Ser Phe Gln Val Ser Leu Thr Gly Phe Le	eu
	Met Leu Glu Ile Val Leu Gly Leu Gly Ser Asn Leu Thr Val Leu Va 50 55 60	1
30	Leu Tyr Cys Met Lys Ser Asn Leu Ile Asn Ser Val Ser Asn Ile Il 65 70 75 80	
	Thr Met Asn Leu His Val Leu Asp Val Ile Ile Cys Val Gly Cys Il	.e
	Pro Leu Thr Ile Val Ile Leu Leu Leu Ser Leu Glu Ser Asn Thr Al	.a

		٠.	•		100	٠.		٠.		105		•		٠.,	110	• .	
		Leu	Ile	Cys 115	Cys	Phe	His	Glu	Ala 120	Cys	Val	Ser	Phe	Ala 125	Ser	Val	Ser
5		Thr	Ala 130	Ile	Asn	Val	Phe	Ala 135	Ile	Thr	Leu	Asp	Arg 140	туг	Asp	Ile	Sei
		Val 145	Lys	Pro	Ala	Asn	Arg 150	Ile	Leu	Thr	Met	Gly 155	Arg	Ala	Val	Met	Let 160
		Met	Ile	Ser	Ile	Trp 165	Ile	Phe	Ser	Phe	Phe 170	Ser	Phe	Leu	Ile	Pro 175	Phe
10	•	Ile	Glu	Val	Asn 180	Phe	Phe	Ser	Leu	Gln 185	Ser	Gly	Asn ,	Thr	Trp 190	Glu	Asr
		Lys	Thr	Leu 195	Leu	Cys	Val	Ser	Thr 200	Asn	Glu	Tyr	Tyr	Thr 205	Glu	Leu	Gl
15		Met	Tyr 210	Tyr	His	Leu	Leu	Val 215	Gln	Ile	Pro	Ile	Phe 220	Phe	Phe	Thr	Val
		Val 225	Val	Met	Leu	Ile	Thr 230	Tyr	Thr	Lys	Ile	Leu 235	Gln	Ala	Leu	Asn	11e 240
		Arg	Ile	Gly	Thr	Arg 245	Phe	Ser	Thr	Gly	Gln 250	Lys	Lys	Lys	Ala	Arg 255	Lys
20		Lys	Lys	Thr	Ile 260	Ser	Leu	Thr	Thr	Gln 265	His	Glu	Ala	Thr	Asp 270	Met	Ser
		Gln	Ser	Ser 275	Gly	Gly	Arg	Asn	Val 280	Val	Phe	Gly	Val	Arg 285	Thr	Ser	Val
25		Ser	Val 290	Ile	Ile	Ala	Leu	Arg 295	Arg	Ala	Val	Lys	Arg 300	His	Arg	Glu	Arg
		Arg 305	Glu	Arg	Gln	Lys	Arg 310	Val	Lys	Arg	Met	Ser 315	Leu	Leu	Ile	Ile	Ser 320
		Thr	Phe	Leu	Leu	Cys 325	Trp	Thr	Pro	Ile	Ser 330	Val	Leu	Asn		Thr 335	Ile
30		Leu	Cys	Leu	Gly 340	Pro	Ser	Asp	Leu	Leu 345	Val	Lys	Leu	Arg	Leu 350	Cys	Phe
		Leu	Val	Met 355	Ala	Tyr	Gly	Thr	Thr 360	Ile	Phe	His	Pro	Leu 365	Leu	Tyr	Ala
35		Phe	Thr 370	Arg	Gln	Lys	Phe	Gln 375	Lys	Val	Leu	Lys	Ser 380	Lys	Met	Lys	Lys
		Arg 385	Val	Val	Ser	Ile	Val 390	Glu	Ala	Asp	Pro	Leu 395	Pro	Asn	Asn	Ala	Val

Ile His Asn Ser Trp Ile Asp Pro Lys Arg Asn Lys Lys Ile Thr Phe
405 410 415

Glu Asp Ser Glu Ile Arg Glu Lys Arg Leu Val Pro Gln Val Val Thr
420 425 430

Asp

(192) INFORMATION FOR SEQ ID NO:191:

(192) INFORMATION FOR BEG 15 NO.131

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1209 base pairs
- 10 (B) TYPE: nucleic acid

5

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:191:

15	ATGTTGTGTC	CTTCCAAGAC	AGATGGCTCA	GGGCACTCTG	GTAGGATTCA	CCAGGAAACT	60
	CATGGAGAAG	GGAAAAGGGA	CAAGATTAGC	AACAGTGAAG	GGAGGGAGAA	TGGTGGGAGA	120
	GGATTCCAGA	TGAACGGTGG	GTCGCTGGAG	GCTGAGCATG	CCAGCAGGAT	GTCAGTTCTC	180
•	AGAGCAAAGC	CCATGTCAAA	CAGCCAACGC	TTGCTCCTTC	TGTCCCCAGG	ATCACCTCCT	240
·	CGCACGGGGA	GCATCTCCTA	CATCAACATC	ATCATGCCTT	CGGTGTTCGG	CACCATCTGC	300
20	CTCCTGGGCA	TCATCGGGAA	CTCCACGGTC	ATCTTCGCGG	TCGTGAAGAA	GTCCAAGCTG	360
	CACTGGTGCA	ACAACGTCCC	CGACATCTTC	ATCATCAACC	TCTCGGTAGT	AGATCTCCTC	420
	TTTCTCCTGG	GCATGCCCTT	CATGATCCAC	CAGCTCATGG	GCAATGGGGT	GTGGCACTTT	480
	GGGGAGACCA	TGTGCACCCT	CATCACGGCC	ATGGATGCCA	ATAGTCAGTT	CACCAGCACC	540
	TACATCCTGA	CCGCCATGGC	CATTGACCGC	TACCTGGCCA	CTGTCCACCC	CATCTCTTCC	600
25	ACGAAGTTCC	GGAAGCCCTC	TGTGGCCACC	CTGGTGATCT	GCCTCCTGTG	GGCCCTCTCC	660
	TTCATCAGCA	TCACCCCTGT	GTGGCTGTAT	GCCAGACTCA	TCCCCTTCCC	AGGAGGTGCA	720
	GTGGGCTGCG	GCATACGCCT	GCCCAACCCA	GACACTGACC	TCTACTGGTT	CACCCTGTAC	780
	CAGTTTTTCC	TGGCCTTTGC	CCTGCCTTTT	GTGGTCATCA	CAGCCGCATA	CGTGAGGATC	840
	CTGCAGCGCA	TGACGTCCTC	AGTGGCCCCC	GCCTCCCAGC	GCAGCATCCG	GCTGCGGACA	900
30	AAGAGGGTGA	AACGCACAGC	CATCGCCATC	TGTCTGGTCT	TCTTTGTGTG	CTGGGCACCC	960
	TACTATGTGC	TACAGCTGAC	CCAGTTGTCC	ATCAGCCGCC	CGACCCTCAC	CTTTGTCTAC	1020
	TTATACAATG	CGGCCATCAG	CTTGGGCTAT	GCCAACAGCT	GCCTCAACCC	CTTTGTGTAC	1080

BNSDOCID -WO COSSISSATI -

145

ATCGTGCTCT GTGAGACGTT CCGCAAACGC TTGGTCCTGT CGGTGAAGCC TGCAGCCCAG 1140

	GGGCAGCTTC GCGCTGTCAG CA	ACGCTCAG ACGGCTG	ACG AGGAGAGGAC AGAAAGCAAA	1200
	GGCACCTGA	•		1209
	(193) INFORMATION FOR SE	Q ID NO:192:		<i>:</i>
5	(i) SEQUENCE CHARA (A) LENGTH: 40 (B) TYPE: amin (C) STRANDEDNE (D) TOPOLOGY:	2 amino acids no acid nss:		·
10	(ii) MOLECULE TYPE	: protein		
	(xi) SEQUENCE DESC	RIPTION: SEQ ID N	JO:192:	
	Met Leu Cys Pro Ser 1 5	Lys Thr Asp Gly	Ser Gly His Ser Gly Arg I	le
15	His Gln Glu Thr His 20	Gly Glu Gly Lys 25	Arg Asp Lys Ile Ser Asn S 30	er
	Glu Gly Arg Glu Asn 35	Gly Gly Arg Gly 40	Phe Gln Met Asn Gly Gly S 45	er -
	Leu Glu Ala Glu His 50	Ala Ser Arg Met 55	Ser Val Leu Arg Ala Lys P 60	ro
20	Met Ser Asn Ser Gln 65	Arg Leu Leu Leu 70	Leu Ser Pro Gly Ser Pro P 75 8	
	Arg Thr Gly Ser Ile 85		Ile Ile Met Pro Ser Val P 90 95	he
25	Gly Thr Ile Cys Leu 100	Leu Gly Ile Ile 105	Gly Asn Ser Thr Val Ile P	he
	Ala Val Val Lys Lys 115	Ser Lys Leu His	Trp Cys Asn Asn Val Pro A 125	sp
	130	135	Asp Leu Leu Phe Leu Leu G 140	
30	Met Pro Phe Met Ile 145	His Gln Leu Met	Gly Asn Gly Val Trp His Pl 155 10	he 60
	Gly Glu Thr Met Cys 165		Ala Met Asp Ala Asn Ser G 170 175	ln
35	Phe Thr Ser Thr Tyr 180	Ile Leu Thr Ala 1	Met Ala Ile Asp Arg Tyr Le 190	eu
	Ala Thr Val His Pro	Ile Ser Ser Thr	Lys Phe Arg Lys Pro Ser Va	al

	.•		195				٠	200			12 L		205	*		
	Ala	Thr 210	Leu	Val	Ile	Cys	Leu 215	Leu	Trp	Ala	Leu	Ser 220	Phe	Ile	Ser	Ile
5	Thr 225	Pro	Val	Trp	Leu	Tyr 230	Ala	Arg	Leu	Ile		Phe	Pro	Gly	Gly	Ala 240
-	Val	Gly	Cys	Gly	Ile 245	Arg	Leu	Pro	Asn	Pro 250	Asp	Thr	Asp	Leu	Tyr 255	Trp
	Phe	Thr	Leu	Tyr 260	Gln	Phe	Phe	Leu	Ala 265	Phe	Ala	Leu	Pro	Phe 270	Val	Val
10	Ile	Thr	Ala 275	Ala	Tyr	Val	Arg	Ile 280	Leu	Gln	Arg	Met	Thr 285	Ser	Ser	Val
	Ala	Pro 290		Ser	Gln	Arg	Ser 295	Ile		Leu	Arg	Thr 300	Lys	Arg	Val	Lys
15	Arg 305	Thr	Ala	Ile	Ala	Ile 310	Cys	Leu	Val	Phe	Phe 315	Val	Cys	Trp	Ala	Pro 320
	Tyr	Tyr	Val	Leu	Gln 325	Leu	Thr	Gln	Leu	Ser 330	Ile	Ser	Arg		Thr 335	Leu
	Thr	Phe	Val	Tyr 340	Leu	Tyr	Asn	Ala	Ala 345	Ile	Ser	Leu	Gly	Tyr 350	Ala	Asn
20	Ser	Cys	Leu 355	Asn	Pro	Phe	Val	Tyr 360	Ile	Val	Leu		Glu 365	Thr	Phe	Arg
	Lys	Arg 370	Leu	Val	Leu	Ser	Val 375	Lys	Pro	Ala	Ala	Gln 380	Gly	Gln	Leu	Arg
25	Ala 385	Val	Ser	Asn	Ala	Gln 390	Thr	Ala	Asp	Glu	Glu 395		Thr	Glu	Ser	Lys 400
	Gly	Thr										-				
	(194) IN	FORM	ATIO1	1 FOF	SEÇ) ID	NO:1	.93 :								
30	(i	(B)	LEN TYP	NGTH: PE: 1 RANDE	112 ucle	28 ba eic a SS: s	ase p acid singl	airs	.	٠,.						
	(i	i) MC	OLECU	TLE I	YPE:	DNA	ı (ge	nomi	.c)							
35	(x	i) SI	EQUEN	ICE I	ESCR	RIPTI	ON:	SEQ	ID N	iO:19	3:					
	ATGGATGT	GA CI	TCCC	CAAGO	cce	GGGC	GTG	GGCC	TGGA	GA 1	GTAC	CCAG	G CA	'CCGC	GCAC	9 60
	GCTGCGGC	cc co	CAACA	CCAC	CTC	cccc	GAG	CTCA	ACCI	GT C	CCAC	CCGC	T CC	TGGG	CACC	120

#4500000 -MU 000010041 1 -

147

				1 17			
	GCCCTGGCCA	ATGGGACAGG	TGAGCTCTCG	GAGCACCAGC	AGTACGTGAT	CGGCCTGTTC	180
	CTCTCGTGCC	TCTACACCAT	CTTCCTCTTC	CCCATCGGCT	TTGTGGGCAA	CATCCTGATC	240
	CTGGTGGTGA	ACATCAGCTT	CCGCGAGAAG	ATGACCATCC	CCGACCTGTA	CTTCATCAAC	300
	CTGGCGGTGG	CGGACCTCAT	CCTGGTGGCC	GACTCCCTCA	TTGAGGTGTT	CAACCTGCAC	360
4	GAGCGGTACT	ACGACATCGC	CGTCCTGTGC	ACCTTCATGT	CGCTCTTCCT	GCAGGTCAAC	420
	ATGTACAGCA	GCGTCTTCTT	CCTCACCTGG	ATGAGCTTCG	ACCGCTACAT	CGCCCTGGCC	480
	AGGGCCATGC	GCTGCAGCCT	GTTCCGCACC	AAGCACCACG	CCCGGCTGAG	CTGTGGCCTC	540
	ATCTGGATGG	CATCCGTGTC	AGCCACGCTG	GTGCCCTTCA	CCGCCGTGCA	CCTGCAGCAC	600
	ACCGACGAGG	CCTGCTTCTG	TTTCGCGGAT	GTCCGGGAGG	TGCAGTGGCT	CGAGGTCACG	660
10) CTGGGCTTCA	TCGTGCCCTT	CGCCATCATC	GGCCTGTGCT	ACTCCCTCAT	TGTCCGGGTG	720
	CTGGTCAGGG	CGCACCGGCA	CCGTGGGCTG	CGGCCCCGGC	GGCAGAAGGC	GAAACGCATG	780
	ATCCTCGCGG	TGGTGCTGGT	CTTCTTCGTC	TGCTGGCTGC	CGGAGAACGT	CTTCATCAGC	840
	GTGCACCTCC	TGCAGCGGAC	GCAGCCTGGG	GCCGCTCCCT	GCAAGCAGTC	TTTCCGCCAT	900
	GCCCACCCCC	TCACGGGCCA	CATTGTCAAC	CTCGCCGCCT	TCTCCAACAG	CTGCCTAAAC	960
	CCCCTCATCT	ACAGCTTTCT	CGGGGAGACC	TTCAGGGACA	AGCTGAGGCT	GTACATTGAG	1020
	CAGAAAACAA	ATTTGCCGGC	CCTGAACCGC	TTCTGTCACG	CTGCCCTGAA	GGCCGTCATT	1080
	CCAGACAGCA	CCGAGCAGTC	GGATGTGAGG	TTCAGCAGTG	CCGTGTGA	٠,	1128
	(195) INFO	RMATION FOR	SEQ ID NO:1	194:			
20	, - ,	SEQUENCE CHA (A) LENGTH:					

- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:194:

Met Asp Val Thr Ser Gln Ala Arg Gly Val Gly Leu Glu Met Tyr Pro 10

Gly Thr Ala His Ala Ala Ala Pro Asn Thr Thr Ser Pro Glu Leu Asn 25

30 Leu Ser His Pro Leu Leu Gly Thr Ala Leu Ala Asn Gly Thr Gly Glu 35 40 45

	Leu	ser 50	GIU	Hls	GIN	GIN	55	vaı	11e	GIÀ	Leu	60	Leu	ser	Cys	Leu
-	Tyr 65	Thr	Ile	Phe	Leu	Phe	Pro	Ile	Gly	Phe	Val 75	Gly	Asn	Ile	Leu	Ile 80
5	Leu	Val	Val	Asn	Ile 85	Ser	Phe	Arg	Glu	Lys 90	Met	Thr	Ile	Pro	Asp 95	Leu
	Tyr	Phe	Ile	Asn 100	Leu	Ala	Val	Ala	Asp 105	Leu	Ile	Leu	Val	Ala 110	Asp	Ser
10	Leu	Ile	Glu 115	Val	Phe	Asn		His 120	Glu	Arg	Tyr	Tyr	Asp 125	Ile	Ala	Val
	Leu	Cys 130	Thr	Phe	Met	Ser	Leu 135		Leu	Gln	Val	Asn 140		Tyr	Ser	Ser
	Val 145	Phe	Phe	Leu	Thr	Trp 150	Met	Ser	Phe	Asp	Arg 155	Tyr	Ile	Ala	Leu	Ala 160
15	Arg	Ala	Met	Arg	Cys 165	Ser	Leu	Phe	Arg	Thr 170	Lys	His	His	Ala	Arg 175	Leu
•	Ser	Cys	Gly	Leu 180	Ile	Trp	Met	Ala	Ser 185	Val	Ser	Ala	Thr	Leu 190	Val	Pro
20	Phe	Thr	Ala 195		His	Leu	Gln	His 200	Thr	Asp	Glu	Ala	Cys 205		Cys	Phe
•	Ala	Asp 210	Val	Arg	Glu	Val	Gln 215	Trp	Leu	Glu	Val	Thr 220	Leu	Gly	Phe	Ile
	Val 225	Pro	Phe	Ala	Ile	Ile 230	Gly	Leu	Cys	Tyr	Ser 235	Leu	Ile	Val	Arg	Val 240
25	Leu	Val	Arg	Ala	His 245	Arg	His	Arg	Gly	Leu 250	Arg	Pro	Arg	Arg	Gln 255	Lys
	Ala	Lys	Arg	Met 260	Ile	Leu	Ala	Val	Val 265	Leu	Val	Phe	Phe	Val 270	Cys	Trp
30	Leu	Pro	Glu 275	Asn	Val	Phe	Ile	Ser 280	Val	His	Leu	Leu	Gln 285	Arg	Thr	Gln
	Pro	Gly 290	Ala	Ala	Pro	Cys	Lys 295	Gln	Ser	Phe	Arg	His 300	Ala	His	Pro	Leu
0.5	305	Gly				310					315					320
35		Leu			325					330					335	
	Len	Tvr	Tle	Glu	Gln	Lvs	Thr	Asn	Len	Pro	Ala	1.611	Asn	Ara	Phe	CVS

149

340 345 His Ala Ala Leu Lys Ala Val Ile Pro Asp Ser Thr Glu Gln Ser Asp 360 Val Arg Phe Ser Ser Ala Val 375 5 370 (196) INFORMATION FOR SEQ ID NO:195: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 960 base pairs (B) TYPE: nucleic acid 10 (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (genomic) (xi) SEQUENCE DESCRIPTION: SEQ ID NO:195: ATGCCATTCC CAAACTGCTC AGCCCCCAGC ACTGTGGTGG CCACAGCTGT GGGTGTCTTG 60 15 CTGGGGCTGG AGTGTGGGCT GGGTCTGCTG GGCAACGCGG TGGCGCTGTG GACCTTCCTG 120 TTCCGGGTCA GGGTGTGGAA GCCGTACGCT GTCTACCTGC TCAACCTGGC CCTGGCTGAC 180 CTGCTGTTGG CTGCGTGCCT GCCTTTCCTG GCCGCCTTCT ACCTGAGCCT CCAGGCTTGG 240 CATCTGGGCC GTGTGGGCTG CTGGGCCCTG CGCTTCCTGC TGGACCTCAG CCGCAGCGTG 300 GGGATGGCCT TCCTGGCCGC CGTGGCTTTG GACCGGTACC TCCGTGTGGT CCACCCTCGG 360 20 CTTAAGGTCA ACCTGCTGTC TCCTCAGGCG GCCCTGGGGG TCTCGGGCCT CGTCTGGCTC 420 CTGATGGTCG CCCTCACCTG CCCGGGCTTG CTCATCTCTG AGGCCGCCCA GAACTCCACC 480 AGGTGCCACA GTTTCTACTC CAGGGCAGAC GGCTCCTTCA GCATCATCTG GCAGGAAGCA 540 CTCTCCTGCC TTCAGTTTGT CCTCCCCTTT GGCCTCATCG TGTTCTGCAA TGCAGGCATC 600 ATCAGGGCTC TCCAGAAAAG ACTCCGGGAG CCTGAGAAAC AGCCCAAGCT TCAGCGGGCC 660 25 AAGGCACTGG TCACCTTGGT GGTGGTGCTG TTTGCTCTGT GCTTTCTGCC CTGCTTCCTG 720 GCCAGAGTCC TGATGCACAT CTTCCAGAAT CTGGGGAGCT GCAGGGCCCT TTGTGCAGTG 780 GCTCATACCT CGGATGTCAC GGGCAGCCTC ACCTACCTGC ACAGTGTCGT CAACCCCGTG 840 GTATACTGCT TCTCCAGCCC CACCTTCAGG AGCTCCTATC GGAGGGTCTT CCACACCCTC 900 CGAGGCAAAG GGCAGGCAGC AGAGCCCCCA GATTTCAACC CCAGAGACTC CTATTCCTGA 960 30 (197) INFORMATION FOR SEQ ID NO:196:

(i) SEQUENCE CHARACTERISTICS:

		(B) TY) ST	PE: RAND	: 31 amin EDNE GY:	o ac SS:	id				. •					
5	(i	i) M	OLEC	ULE	TYPE	: pr	otei	n								
	(x	i) S	EQUE	NCE	DESC	RIPT	ION:	SEQ	ID :	NO:1	96:					
	Met 1	Pro	Phe	Pro	Asn 5	Cys	Ser	Ala	Pro	Ser 10	Thr	Val	Val	Ala	Thr 15	Al
10	Val	Gly	Val	Leu 20	Leu	Gly	Leu	Glu	Cys 25	Gly	Leu	Gly	Leu	Leu 30	Gly	As
	Ala	Val	Ala 35	Leu	Trp	Thr	Phe	Leu 40	Phe	Arg	Val	Arg	Val 45	Trp	Lys	Pr
	Tyr	Ala 50	Val	Tyr	Leu	Leu	Asn 55	Leu	Ala	Leu	Ala	Asp 60	Leu	Leu	Leu	Ala
15	Ala 65	Cys	Leu	Pro	Phe	Leu 70	Ala	Ala	Phe	Tyr	Leu 75	Ser	Leu	Gln	Ala	Tr) 80
	His	Leu	Gly	Arg	Val 85	Gly	Cys	Trp	Ala	Leu 90	Arg	Phe	Leu	Leu	Asp 95	Let
20	Ser	Arg	Ser	Val 100	Gly	Met	Ala	Phe	Leu 105	Ala	Ala	Val	Ala	Leu 110	Asp	Arg
	Tyr	Leu	Arg 115	Val	Val	His	Pro	Arg 120	Leu	Lys	Val	Asn	Leu 125	Leu	Ser	Pro
	Gln	Ala 130	Ala	Leu	Gly	Val	Ser 135	Gly	Leu	Val	Trp	Leu 140	Leu	Met	Val	Ala
25	Leu 145	Thr	Cys	Pro	Gly	Leu 150	Leu	Ile	Ser	Glu	Ala 155	Ala	Gln	Asn	Ser	Th:
	Arg	Cys	His	Ser	Phe 165	Tyr	Ser	Arg	Ala	Asp 170	Gly	Ser	Phe	Ser	Ile 175	Ile
30	Trp	Gln	Glu	Ala 180	Leu	Ser	Cys	Leu	Gln 185	Phe	Val	Leu	Pro	Phe 190	Gly	Let
	Ile	Val	Phe 195	Cys	Asn	Ala	Gly	Ile 200	Ile	Arg	Ala	Leu	Gln 205	Lys	Arg	Leu
٠	Arg	Glu 210	Pro	Glu	Lys	Gln	Pro 215	Lys	Leu	Gln	Arg	Ala 220	Lys	Ala	Leu	Va]
35	Thr 225	Leu	Val	Val	Val	Leu 230	Phe	Ala	Leu	Cys	Phe 235	Leu	Pro	Cys	Phe	Let 240
	Ala	Arg	Val	Leu	Met	His	Ile	Phe	Gln	Asn	Lei	Glv	Ser	Cve	Ara	- ומ

		2	245 .		250	• •	255	
	Leu Cy	ys Ala Val 2 260	Ala His T	Thr Ser	Asp Val 265	Thr Gly	Ser Leu Thr 270	Tyr
5	Leu Hi	is Ser Val V 275	/al Asn P	Pro Val 280	Val Tyr	-	Ser Ser Pro 285	Thr
		rg Ser Ser : 90	_	arg Val	Phe His	Thr Leu 2	Arg Gly Lys	Gly
	Gln Al	la Ala Glu I	Pro Pro A 310	asp Phe	Asn Pro	Arg Asp :	Ser Tyr Ser	
10	(198) INFOR	RMATION FOR	SEQ ID N	10:197:				
15	(SEQUENCE CHA (A) LENGTH: (B) TYPE: no (C) STRANDEI (D) TOPOLOGY	1143 bas acleic ac ONESS: si	se pairs aid .ngle				
	(ii)	MOLECULE TY	PE: DNA	(genomi	c)			5
	(xi)	SEQUENCE DE	ESCRIPTIO	N: SEQ	ID NO:1	97:		3
	ATGGAGGAAG	GTGGTGATTT	TGACAACT	AC TATG	GGGCAG	ACAACCAGT	C TGAGTGTGAC	60
	TACACAGACT	GGAAATCCTC	GGGGGCCC	TC ATCC	CTGCCA	TCTACATGT	r ggtcttcctc	120
20	CTGGGCACCA	CGGGAAACGG	TCTGGTGC	TC TGGA	CCGTGT '	TTCGGAGCA	G CCGGGAGAA	3 180
	AGGCGCTCAG	CTGATATCTT	CATTGCTA	GC CTGG	CGGTGG	CTGACCTGA	C CTTCGTGGT	240
	ACGCTGCCCC	TGTGGGCTAC	CTACACGT	CAC CGGG	ACTATG	ACTGGCCCT	TGGGACCTT	300
	TTCTGCAAGC	TCAGCAGCTA	CCTCATCT	TC GTCA	ACATGT	ACGCCAGCG'	r cttctgcctc	360
	ACCGGCCTCA	GCTTCGACCG	CTACCTGG	SCC ATCG	TGAGGC	CAGTGGCCA	A TGCTCGGCTC	420
25	AGGCTGCGGG	TCAGCGGGGC	CGTGGCCA	CG GCAG	TTCTTT	GGGTGCTGG	C CGCCCTCCTC	480
	GCCATGCCTG	TCATGGTGTT	ACGCACCA	CC GGGG	ACTTGG .	AGAACACCA	C TAAGGTGCAG	540
	TGCTACATGG	ACTACTCCAT	GGTGGCCA	ACT GTGA	GCTCAG	AGTGGGCCT	G GGAGGTGGG	600
	CTTGGGGTCT	CGTCCACCAC	CGTGGGCT	TT GTGG	TGCCCT	TCACCATCA	r GCTGACCTGT	r 660
	TACTTCTTCA	TCGCCCAAAC	CATCGCTG	GC CACT	TCCGCA .	AGGAACGCA'	r CGAGGGCCTC	720
30	CGGAAGCGGC	GCCGGCTTAA	GAGCATCA	ATC GTGG	TGCTGG	TGGTGACCT	r TGCCCTGTG	780
	TGGATGCCCT	ACCACCTGGT	GAAGACGC	TG TACA	TGCTGG	GCAGCCTGC	T GCACTGGCC	840
	TGTGACTTTG	ACCTCTTCCT	CATGAACA	ATC TTCC	CCTACT	GCACCTGCA	r CAGCTACGTO	900

	AACAGCT	GCC 7	CAAC	ccc	rt .co	CTCT	ATGCC	TT	TTC	SACC	CCC	CTT	CCG (CCAG	CCT	GC.	960
	ACCTCCA	TGC 1	rctgo	CTGT	G CC	CAGAC	CAGG	TGC	GCAC	GCA	CCT	CCAC	CAG (CAGC	GTGG	G 1	.020
	GAGAAGT	CAG C	CCAGO	TACI	rc Ti	CGGC	GCAC	AGC	CAGG	GGC	CCG	cccc	CAA C	CATCO	GCAA	.G 1	.080
	GGTGGAG	AAC A	GATG	CACC	SA GA	AATO	CATO	ccc	TACA	/GCC	AGG	GACC	CT 1	rgrge	TTGA	C 1	140
5	TAG															1	143
	(199) II	1FORM	IATIC	N FC	R SE	Q II	NO:	198:									
10	(:	(E	L) LE 3) TY 3) ST	NGTH PE: RAND	: 38 amin EDNE	0 am o ac SS:	ino	acid						: .			
	,, . (ii) M	OLEC	ULE	TYPE	: pr	otei	n - , ·				٠.٠		٠٠			
	()	(i) S	EQUE	NCE	DESC	RIPT	'ION:	SEQ	ID	NO:1	98:						
15	Met 1	Glu	Glu	Gly	Gly 5	Asp	Phe	Asp	Asn	Tyr 10	Tyr	Gly	Ala	Asp	Asn 15	Gln	
•	Ser	Glu	Cys	Glu 20	Tyr	Thr	Asp	Trp	Lys 25		Ser	Gly	Ala	Leu 30	Ile	Pro	
	Ala	lle	Туг 35	Met	Leu	Val	Phe	Leu 40	Leu	Gly	Thr	Thr	Gly 45	Asn	Gly	Leu	
20	Val	Leu 50	Trp	Thr	Val	Phe	Arg 55	Ser	Ser	Arg	Glu	Lys 60	Arg	Arg	Ser	Ala	
	Asp 65	Ile	Phe	Ile	Ala	Ser 70	Leu	Ala	Val	Ala	Asp 75	Leu	Thr	Phe	Val	Val 80	
25	Thr	Leu	Pro	Leu	Trp 85	Ala	Thr	Tyr	Thr	Tyr 90	Arg	Asp	Tyr	Asp	Trp 95	Pro	
	Phe	Gly	Thr	Phe 100	Phe	Cys	Lys	Leu	Ser 105	Ser	Tyr	Leu	Ile	Phe 110	Val	Asn	
	Met	Tyr	Ala 115	Ser	Val	Phe	Cys	Leu 120		Gly	Leu	Ser	Phe 125	Asp	Arg	Tyr	
30	Leu	Ala 130	Ile	Val	Arg	Pro	Val 135	Ala	Asn	Ala	Arg	Leu 140	Arg	Leu	Arg	Val	
	Ser 145	Gly	Ala	Val	Ala	Thr 150	Ala	Val	Leu	Trp	Val 155	Leu	Ala	Ala	Leu	Leu 160	
35	Ala	Met	Pro	Val	Met 165	Val	Leu	Arg	Thr	Thr 170	Gly	Asp	Leu	Glu	Asn 175	Thr	

		Thr	Lys	Val	Gln 180	Cys	Tyr	Met	Asp	Tyr 185	Ser	Met	Val	Ala	Thr 190	Val	Ser	
		Ser	Glu	Trp 195	Ala	Trp	Glu	Val	Gly 200	Leu	Gly	Val	Ser	Ser 205	Thr	Thr	Val	
5		Gly	Phe 210	Val	Val	Pro	Phe	Thr 215	Ile	Met	Leu	Thr	Cys 220	Tyr	Phe	Phe	Ile	
		Ala 225	Gln	Thr	Ile	Ala	Gly 230	His	Phe	Arg	Lys	Glu 235	Arg	Ile	Glu	Gly	Leu 240	
10		Arg	Lys	Arg	Arg	Arg 245	Leu	Lys	Ser	Ile	Ile 250	Val	Val	Leu	Val	Val 255	Thr	
		Phe	Ala	Leu	Cys 260	Trp	Met	Pro	Tyr	His 265	Leu	Val	Lys	Thr	Leu 270	Tyr	Met	
		Leu	Gly	Ser 275	Leu	Leu	His	Trp	Pro 280	Cys	Asp	Phe .	Asp	Leu 285	Phe	Leu	Met	
15		Asn	Ile 290	Phe	Pro	Tyr	Cys	Thr 295	Cys	Ile	Ser	Tyr	Val 300	Asn	Ser	Cys	Leu	
		Asn 305	Pro	Phe	Leu	Tyr	Ala 310	Phe	Phe	Asp	Pro	Arg 315	Phe	Arg	Gln	Ala	Cys 320	
20		Thr	Ser	Met		Cys 325		Gly	Gln		Arg 330	Cys	Ala	Gly	Thr	Ser 335	His	,
		Ser	Ser	Ser	Gly 340	Glu	Lys	Ser	Ala	Ser 345	Tyr	Ser	Ser	Gly	His 350	Ser	Gln	
		Gly	Pro	Gly 355	Pro	Asn	Met	Gly	Lys 360	Gly	Gly	Glu	Gln	Met 365	His	Glu	Lys	
25		Ser	Ile 370	Pro	Tyr	Ser	Gln	Glu 375	Thr	Leu	Val	Val	Asp 380					
	(200)	INF	FORM	MOITA	FOF	SEÇ	Q ID	NO: 1	199:						-			
30	•	(i)	(A) (B) (C)	LEN TYI STR	IGTH: PE: r RANDE	HARAC : 113 nucle EDNES EY: 3	l9 ba eic a SS: s	ase p acid singl	oairs	3		•						
						TYPE:		_			IO . 1 (
35	ATGAA					DESCI							במאכי	TH (18	ייייטיי	2007 7	.	60
رر													*					
	CTGGA	*CMG	71 I.C	OMU	7WC 13	2 IM	-CGM	ACC	1000	. 1	4 90.		· CMIC	- 1 C I	بحدد	-100	- 4	. 2 0

	ACAGAGGGTC	CCCTCATGGC	CTCCTTCAAG	GCCGTGTTCG	TGCCCGTGGC	CTACAGCCTC	180
	ATCTTCCTCC	TGGGCGTGAT	CGGCAACGTC	CTGGTGCTGG	TGATCCTGGA	GCGGCACCGG	240
	CAGACACGCA	GTTCCACGGA	GACCTTCCTG	TTCCACCTGG	CCGTGGCCGA	CCTCCTGCTG	300
	GTCTTCATCT	TGCCCTTTGC	CGTGGCCGAG	GGCTCTGTGG	GCTGGGTCCT	GGGGACCTTC	360
5	CTCTGCAAAA	CTGTGATTGC	CCTGCACAAA	GTCAACTTCT	ACTGCAGCAG	CCTGCTCCTG	420
	GCCTGCATCG	CCGTGGACCG	CTACCTGGCC	ATTGTCCACG	CCGTCCATGC	CTACCGCCAC	480
	CGCCGCCTCC	TCTCCATCCA	CATCACCTGT	GGGACCATCT	GGCTGGTGGG	CTTCCTCCTT	540
	GCCTTGCCAG	AGATTCTCTT	CGCCAAAGTC	AGCCAAGGCC	ATCACAACAA	CTCCCTGCCA	600
	CGTTGCACCT	TCTCCCAAGA	GAACCAAGCA	GAAACGCATG	CCTGGTTCAC	CTCCCGATTC	660
10	CTCTACCATG	TGGCGGGATT	CCTGCTGCCC	ATGCTGGTGA	TGGGCTGGTG	CTACGTGGGG	720
	GTAGTGCACA	GGTTGCGCCA	GGCCCAGCGG	CGCCCTCAGC	GGCAGAAGGC	AAAAAGGGTG	780
	GCCATCCTGG	TGACAAGCAT	CTTCTTCCTC	TGCTGGTCAC	CCTACCACAT	CGTCATCTTC	840
	CTGGACACCC	TGGCGAGGCT	GAAGGCCGTG	GACAATACCT	GCAAGCTGAA	TGGCTCTCTC	900
	CCCGTGGCCA	TCACCATGTG	TGAGTTCCTG	GGCCTGGCCC	ACTGCTGCCT	CAACCCCATG	960
15	CTCTACACTT	TCGCCGGCGT	GAAGTTCCGC	AGTGACCTGT	CGCGGCTCCT	GACCAAGCTG	1020
	GGCTGTACCG	GCCCTGCCTC	CCTGTGCCAG	CTCTTCCCTA	GCTGGCGCAG	GAGCAGTCTC	1080
	TCTGAGTCAG	AGAATGCCAC	CTCTCTCACC	ACGTTCTAG			1119
	(201) TNTPOT	MARTON FOR	CEO TO NO.	200.			

(201) INFORMATION FOR SEQ ID NO:200:

(i) SEQUENCE CHARACTERISTICS: 20

(A) LENGTH: 372 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:200:

Met Asn Tyr Pro Leu Thr Leu Glu Met Asp Leu Glu Asn Leu Glu Asp 5

Leu Phe Trp Glu Leu Asp Arg Leu Asp Asn Tyr Asn Asp Thr Ser Leu

30 Val Glu Asn His Leu Cys Pro Ala Thr Glu Gly Pro Leu Met Ala Ser 40

· Phe Lys Ala Val Phe Val Pro Val Ala Tyr Ser Leu Ile Phe Leu Leu

			50	i			.•	55		**		٠.	60				
		Gly 65	Val	Ile	Gly	Asn	Val 70	Leu	Val	Leu	Val	Ile 75	Leu	Glu	Arg	His	Arg 80
5		Gln	Thr	Arg	Ser	Ser 85	Thr	Glu	Thr	Phe	Leu 90	Phe	His	Leu	Ala	Val 95	Ala
		Asp	Leu	Leu	Leu 100	Val	Phe	Ile	Leu	Pro 105	Phe	Ala	Val	Ala	Glu 110	Gly	Ser
	-	Val	Gly	Trp 115	Val	Leu	Gly	Thr	Phe 120	Leu	Cys	Lys	Thr	Val 125	Ile	Ala	Leu
10		His	Lys 130	Val	Asn	Phe	Tyr	Cys 135	Ser	Ser	Leu	Leu	Leu 140	Ala	Cys	Ile	Ala
		Val 145	Asp	Arg	Tyr	Leu	Ala 150	Ile	Val	His	Ala	Val 155	His	Ala	Tyr	Arg	His 160
15		Arg	Arg	Leu	Leu	Ser 165	Ile	His	Ile	Thr	Cys 170	Gly	Thr	Ile	Trp	Leu 175	Val
		Gly	Phe	Leu	Leu 180	Ala	Leu	Pro	Glu	Ile 185	Leu	Phe	Ala	Lys	Val 190	Ser	Gln
		Gly	His	His 195	Asn	Asn	Ser	Leu	Pro 200	Arg	Cys	Thr	Phe	Ser 205	Gln	Glu	Asn
20	•	Gln	Ala 210	Glu	Thr	His	Ala	Trp 215	Phe	Thr	Ser	Arg	Phe 220	Leu	Tyr	His	Val
		Ala 225	Gly	Phe	Leu	Leu	Pro 230	Met	Leu	Val	Met	Gly 235	Trp	Cys	Tyr	Val	Gly 240
25		Val	Val	His	Arg	Leu 245	Arg	Gln	Ala	Gln	Arg 250	Arg	Pro	Gln	Arg	Gln 255	Lys
		Ala	Lys	Arg	Val 260	Ala	Ile	Leu	Val	Thr 265	Ser	Ile	Phe	Phe	Leu 270	Cys	Trp
		Ser	Pro	Tyr 275	His	Ile	Val	Ile	Phe 280	Leu	Asp	Thr	Leu	Ala 285	Arg	Leu	Lys
30		Ala	Val 290	Asp	Asn	Thr	Cys	Lys 295	Leu	Asn	Gly	Ser	Leu 300	Pro	Val	Ala	Ile
		Thr 305	Met	Cys	Glu	Phe	Leu 310	Gly	Leu	Ala	His	Cys 315	Cys	Leu	Asn	Pro	Met 320
35	-	Leu	Tyr	Thr	Phe	Ala 325	Gly	Val	Lys	Phe	Arg 330	Ser	Asp	Leu	Ser	Arg 335	Leu
	_	Leu	Thr	Lys	Leu 340	Gly	Cys	Thr	Gly	Pro 345	Ala	Ser	Leu	Cys	Gln 350	Leu	Phe

Pro Ser Trp Arg Arg Ser Ser Leu Ser Glu Ser Glu Asn Ala Thr Ser 355 360 365

Leu Thr Thr Phe 370

5 (202) INFORMATION FOR SEQ ID NO:201:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1128 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
- 10 (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:201:

	ATGGATGTGA	CTTCCCAAGO	CCGGGGCGTG	GGCCTGGAGA	TGTACCCAGG	CACCGCGCAG	60
	CCTGCGGCCC	CCAACACCAC	CTCCCCGAG	CTCAACCTGT	CCCACCCGCT	CCTGGGCACC	120
15	GCCCTGGCCA	ATGGGACAGG	TGAGCTCTCG	GAGCACCAGC	AGTACGTGAT	CGGCCTGTTC	180
	CTCTCGTGCC	TCTACACCAT	CTTCCTCTTC	CCCATCGGCT	TTGTGGGCAA	CATCCTGATC	240
	CTGGTGGTGA	ACATCAGCTT	CCGCGAGAAG	ATGACCATCC	CCGACCTGTA	CTTCATCAAC	300
	CTGGCGGTGG	CGGACCTCAT	CCTGGTGGCC	GACTCCCTCA	TTGAGGTGTT	CAACCTGCAC	360
÷,	GAGCGGTACT	ACGACATCGC	CGTCCTGTGC	ACCTTCATGT	CGCTCTTCCT	GCAGGTCAAC	420
20	ATGTACAGCA	GCGTCTTCTT	CCTCACCTGG	ATGAGCTTCG	ACCGCTACAT	CGCCCTGGCC	480
	AGGGCCATGC	GCTGCAGCCT	GTTCCGCACC	AAGCACCACG	CCCGGCTGAG	CTGTGGCCTC	540
	ATCTGGATGG	CATCCGTGTC	AGCCACGCTG	GTGCCCTTCA	CCGCCGTGCA	CCTGCAGCAC	600
	ACCGACGAGG	CCTGCTTCTG	TTTCGCGGAT	GTCCGGGAGG	TGCAGTGGCT	CGAGGTCACG	660
	CTGGGCTTCA	TCGTGCCCTT	CGCCATCATC	GGCCTGTGCT	ACTCCCTCAT	TGTCCGGGTG	720
25	CTGGTCAGGG	CGCACCGGCA	CCGTGGGCTG	CGGCCCCGGC	GGCAGAAGGC	GAAGCGCATG	780
,	ATCCTCGCGG	TGGTGCTGGT	CTTCTTCGTC	TGCTGGCTGC	CGGAGAACGT	CTTCATCAGC	840
	GTGCACCTCC	TGCAGCGGAC	GCAGCCTGGG	GCCGCTCCCT	GCAAGCAGTC	TTTCCGCCAT	900
	GCCCACCCCC	TCACGGGCCA	CATTGTCAAC	CTCACCGCCT	TCTCCAACAG	CTGCCTAAAC	960
	CCCCTCATCT	ACAGCTTTCT	CGGGGAGACC	TTCAGGGACA	AGCTGAGGCT	GTACATTGAG	1020
30	CAGAAAACAA	ATTTGCCGGC	CCTGAACCGC	TTCTGTCACG	CTGCCCTGAA	GGCCGTCATT	1080
	CCAGACAGCA	CCGAÇCAGTC	GGATGTGAGG	TTCAGCAGTG	CCGTGTAG		1128

Committee Service Land Committee

5

*** *

(203) INFORMATION FOR SEQ ID NO:202:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 375 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:202:

		(xi	i) SI	EQUE	ICE I	DESCI	RIPTI	ON:	SEQ	ID 1	10:20	02:					
10		Met 1	Asp	Val	Thr	Ser 5	Gln	Ala	Arg	Gly	Val 10	Gly	Leu	Glu	Met	Tyr 15	Pro
		Gly	Thr	Ala	Gln 20	Pro	Ala	Ala	Pro	Asn 25	Thr	Thr	Ser	Pro	Glu 30	Leu	Asn
		Leu	Ser	His 35	Pro	Leu	Leu	Gly	Thr 40	Ala	Leu	Ala	Asn	Gly 45	Thr	Gly	Glu
15	**	Leu	Ser 50	Glu	His	Gln	Gln	Tyr 55	Val	Ile	Gly	Leu	Phe 60	Leu	Ser	Cys	Leu
		Tyr 65	Thr	Ile	Phe	Leu	Phe 70	Pro	Ile	Gly	Phe	Val 75	Gly	Asn	Ile	Leu	Ile 80
20		Leu	Val	Val	Asn	Ile 85	Ser	Phe	Arg	Glu	Lys 90	Met	Thr	Ile	Pro	Asp 95	Leu
	٠	Tyr	Phe	Ile	Asn 100	Leu	Ala	Val	Ala	Asp 105	Leu	Ile	Leu	Val	Ala 110	Asp	Ser
		Leu	Ile	Glu 115	Val	Phe	Asn	Leu	His 120	Glu	Arg	Tyr	туг	Asp 125	Ile	Ala	Val
25		Leu	Cys 130	Thr	Phe	Met	Ser	Leu 135	Phe	Leu	Gln	Val	Asn 140	Met	Tyr	Ser	Ser
		Val 145	Phe	Phe	Leu	Thr	Trp 150	Met	Ser	Phe	Asp	Arg 155	Tyr	Ile	Ala	Leu	Ala 160
30		Arg	Ala	Met	Arg	Cys 165	Ser	Leu	Phe	Arg	Thr 170	Lys	His	His	Ala	Arg 175	Leu
		Ser	Cys	Gly	Leu 180	Ile	Trp	Met	Ala	Ser 185	Val	Ser	Ala	Thr	Leu 190	Val	Pro
		Phe	Thr	Ala 195	Val	His	Leu	Gln	His 200	Thr	Asp	Glu	Ala	Суs 205	Phe	Cys	Phe
35	•	Ala	Asp 210		Arg	Glu	Val	Gln 215	Trp	Leu	Glu	Val	Thr 220	Leu	Gly	Phe	Ile

Val Pro Phe Ala Ile Ile Gly Leu Cys Tyr Ser Leu Ile Val Arg Val

	225			•		230					235	,		• • •		240	
	Leu	Val	Arg	Ala	His 245	Arg	His	Arg	Gly	Leu 250	Arg	Pro	Arg	Arg	Gln 255	Lys	
5	Ala	Lys	Arg	Met 260	Ile	Leu	Ala	Val	Val 265	Leu	Val	Phe	Phe	Val 270	Cys	Trp	
	Leu	Pro	Glu 275	Asn	Val	Phe	Ile	Ser 280	Val	His	Leu	Leu	Gln 285	Arg	Thr	Gln	
		Gly 290		Ala	Pro		Lys 295	Gln	Ser		Arg		Ala	His	Pro	Leu	
10	Thr 305	Gly	His	Ile	Val	Asn 310	Leu	Thr	Ala	Phe	Ser 315	Asn	Ser	Cys	Leu	Asn 320	
	Pro	Leu		Tyr			Leu					_	_	Lys		Arg	
15	Leu	Tyr	Ile	Glu 340	Gln	Lys	Thr	Asn	Leu 345	Pro	Ala	Leu	Asn	Arg 350	Phe	Cys	
	His	Ala	Ala 355	Leu	Lys	Ala	Val	Ile 360	Pro	Asp	Ser	Thr	Glu 365		Ser	Asp	
	Val	Arg 370		Ser	Ser	Ala	Val 375				•						
20	(204) IN	FORM	ATIO	N FOR	R SE	Q ID	NO:	203:									
25	(i	(B (C) LEI) TY!) ST!	CE CI NGTH: PE: 1 RANDI POLO	: 11: nucle EDNE:	37 ba eic a SS: a	ase pacid	pair	5	•							
		•		٠		• •			: \		٠,						
		i) M							-								
		:i) S															
	ATGGACCT	GG G	GAAA	CCAA:	r gaz	AAAG	CGTG	CTG	GTGG:	rgg (CTCT	CCTT	GT C	ATTT'	rccac	3	60
	GTATGCCT	GT G	TCAA	GATG/	A GG	rcac(GGAC	GAT"	raca:	rcg (GAGA	CAAC	AC C	ACAG:	rgga	2 1	120
30 .	TACACTTT	GT T	CGAG'	TCTT:	r GT	GCTC	CAAG	AAG	GACG:	rgc (GGAA(CTTT	AA AA	GCCT	GTT	2 1	18
	CTCCCTAT	CA T	GTAC'	TCCA'	r ca	TTTG'	TTTC		GCC.	TAC '	TGGG	CAAT	GG G	CTGG	rcgr	3 2	24(
	TTGACCTA	TA T	CTAT	TTCA	A GA	GGCT	CAAG			CCG 2	ATAC	CTAC	ot G	CTCA	ACCTO	3 3	30
	GCGGTGGC	AG A	CATC	CTCT	r cc	rccr	GACC	CTT	CCCT	rct (GGC	CTAC	AG C	GCGG	CCAA	3 3	36
	TCCTGGGT	CT T	CGGT	GTCC	A CT	TTTG	CAAG	CTC	ATCT:	TTG (CCAT	CTAC	AA G	ATGA	GCTT(2 4	42

159

	TTCAGTGGCA TGCTCCTACT TCTTTGCATC AGCATTGACC GCTACGTGGC CATCGTCCAG	480
	GCTGTCTCAG CTCACCGCCA CCGTGCCCGC GTCCTTCTCA TCAGCAAGCT GTCCTGTGTG	540
	GGCATCTGGA TACTAGCCAC AGTGCTCTCC ATCCCAGAGC TCCTGTACAG TGACCTCCAG	600
	AGGAGCAGCA GTGAGCAAGC GATGCGATGC TCTCTCATCA CAGAGCATGT GGAGGCCTTT	660
5	ATCACCATCC AGGTGGCCCA GATGGTGATC GGCTTTCTGG TCCCCCTGCT GGCCATGAGC	720
	TTCTGTTACC TTGTCATCAT CCGCACCCTG CTCCAGGCAC GCAACTTTGA GCGCAACAAG	780
	GCCAAAAAGG TGATCATCGC TGTGGTCGTG GTCTTCATAG TCTTCCAGCT GCCCTACAAT	840
	GGGGTGGTCC TGGCCCAGAC GGTGGCCAAC TTCAACATCA CCAGTAGCAC CTGTGAGCTC	900
	AGTAAGCAAC TCAACATCGC CTACGACGTC ACCTACAGCC TGGCCTGCGT CCGCTGCTGC	960
10	GTCAACCCTT TCTTGTACGC CTTCATCGGC GTCAAGTTCC GCAACGATCT CTTCAAGCTC 1	020
	TTCAAGGACC TGGGCTGCCT CAGCCAGGAG CAGCTCCGGC AGTGGTCTTC CTGTCGGCAC 1	080
	ATCCGGCGCT CCTCCATGAG TGTGGAGGCC GAGACCACCA CCACCTTCTC CCCATAG	137
	(205) INFORMATION FOR SEQ ID NO:204:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 378 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: (D) TOPOLOGY: not relevant 	
	(ii) MOLECULE TYPE: protein	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:204:	
	Met Asp Leu Gly Lys Pro Met Lys Ser Val Leu Val Val Ala Leu Le 1 5 10	1
	Val Ile Phe Gln Val Cys Leu Cys Gln Asp Glu Val Thr Asp Asp Ty 20 25 30	r
25	Ile Gly Asp Asn Thr Thr Val Asp Tyr Thr Leu Phe Glu Ser Leu Cy 35 40 45	5
	Ser Lys Lys Asp Val Arg Asn Phe Lys Ala Trp Phe Leu Pro Ile Me 50 55 60	t
30	Tyr Ser Ile Ile Cys Phe Val Gly Leu Leu Gly Asn Gly Leu Val Va 65 70 75 80	1
	Tour mar The Tur Phe Lys Arg Leu Lys Thr Met Thr Asp Thr Ty	~

Leu Leu Asn Leu Ala Val Ala Asp Ile Leu Phe Leu Leu Thr Leu Pro

					100				, ,	105					110		
		Phe	Trp	Ala 115	Tyr	Ser	Ala	Ala	Lys 120		Trp	Val	Phe	Gly 125	Val	His	Phe
5		Cys	Lys 130	Leu	Ile	Phe	Ala	Ile 135		Lys	Met	Ser	Phe 140	Phe	Ser	Gly	Met
		Leu 145	Leu	Leu	Leu	Cys	Ile 150	Ser	Ile	Asp	Arg	Tyr 155	Val	Ala	Ile	Val	Gln 160
		Ala	Val		Ala		Arg	His	Arg	Ala	Arg 170	Val	Leu	Leu	Ile	Ser 175	Lys
10		Leu	Ser	Cys	Val 180	Gly	Ile	Trp	Ile	Leu 185	Ala	Thr	Val	Leu	Ser 190	Ile	Pro
• .		Glu	Leu	Leu 195	Tyr	Ser	Asp	Leu	Gln 200	Arg	Ser	Ser	Ser	Glu 205	Gln	Ala	Met
15		Arg	Cys 210	Ser	Leu	Ile	Thr	Glu 215	His	Val	Glu	Ala	Phe 220	Ile	Thr	Ile	Gln
		Val 225	Ala	Gln	Met	Val	Ile 230	Gly	Phe	Leu	Val	Pro 235	Leu	Leu	Ala	Met	Ser 240
			Cys			245					250					255	
20			Arg		260					265					270		•
			Val	275					280					285		: .	
25		• .	Asn 290	•	-			295		•			300			٠.	
		305	Ile				310					315					320
20			Asn			325				٠	330		,			335	
30			Phe		340					345					350		
			Gln	355					360	•		Arg	Ser	Ser 365	Met	Ser	Val
35	1055		Ala 370					375		Ser	Pro						
	(206)	INF	ORMA	TION	FOR	SEC	ID	NO:2	:05:								

PCT/US99/23938

5

30

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1086 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:205:

ATGGATATAC AAATGGCAAA CAATTTTACT CCGCCCTCTG CAACTCCTCA GGGAAATGAC 60 TGTGACCTCT ATGCACATCA CAGCACGGCC AGGATAGTAA TGCCTCTGCA TTACAGCCTC GTCTTCATCA TTGGGCTCGT GGGAAACTTA CTAGCCTTGG TCGTCATTGT TCAAAACAGG 10 180 AAAAAATCA ACTCTACCAC CCTCTATTCA ACAAATTTGG TGATTTCTGA TATACTTTTT 240 ACCACGGCTT TGCCTACACG AATAGCCTAC TATGCAATGG GCTTTGACTG GAGAATCGGA 300 GATGCCTTGT GTAGGATAAC TGCGCTAGTG TTTTACATCA ACACATATGC AGGTGTGAAC 360 TTTATGACCT GCCTGAGTAT TGACCGCTTC ATTGCTGTGG TGCACCCTCT ACGCTACAAC 420 AAGATAAAAA GGATTGAACA TGCAAAAGGC GTGTGCATAT TTGTCTGGAT TCTAGTATTT 480 GCTCAGACAC TCCCACTCCT CATCAACCCT ATGTCAAAGC AGGAGGCTGA AAGGATTACA TGCATGGAGT ATCCAAACTT TGAAGAAACT AAATCTCTTC CCTGGATTCT GCTTGGGGCA 600 TGTTTCATAG GATATGTACT TCCACTTATA ATCATTCTCA TCTGCTATTC TCAGATCTGC TGCAAACTCT TCAGAACTGC CAAACAAAAC CCACTCACTG AGAAATCTGG TGTAAACAAA 720 AAGGCTAAAA ACACAATTAT TCTTATTATT GTTGTGTTTG TTCTCTGTTT CACACCTTAC 780 CATGTTGCAA TTATTCAACA TATGATTAAG AAGCTTCGTT TCTCTAATTT CCTGGAATGT 840 AGCCAAAGAC ATTCGTTCCA GATTTCTCTG CACTTTACAG TATGCCTGAT GAACTTCAAT 900 TGCTGCATGG ACCCTTTTAT CTACTTCTTT GCATGTAAAG GGTATAAGAG AAAGGTTATG 960 AGGATGCTGA AACGGCAAGT CAGTGTATCG ATTTCTAGTG CTGTGAAGTC AGCCCCTGAA 1020 GAAAATTCAC GTGAAATGAC AGAAACGCAG ATGATGATAC ATTCCAAGTC TTCAAATGGA 1080 25 1086 AAGTGA

(207) INFORMATION FOR SEQ ID NO:206:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 361 amino acids
- (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant

•	(i:	i) MO	OLECU	JLE 1	YPE:	pro	otein	1.		· 5		٠.	٠.	٠.		
	(x:	i) SI	EQUEN	ICE I	DESC	RIPTI	ON:	SEQ	ID 1	10:20	6:					
	Met 1	Asp	Ile	Gln	Met 5	Ala	Asn	Asn	Phe	Thr 10	Pro	Pro	Ser	Ala	Thr 15	Pro
5	Gln	Gly	Asn	Asp 20	Cys	Asp	Leu	Tyr	Ala 25	His	His	Ser	Thr	Ala 30	Arg	Ile
	Val	Met	Pro 35	Leu	His	Tyr	Ser	Leu 40	Val	Phe	Ile	Ile	Gly 45	Leu	Val	Gly
10	Asn	Leu 50	Leu	Ala	Leu	Val	Val 55	Ile	Val	Gln	Asn	Arg 60	Lys	Lys	Ile	Asn
	Ser 65	Thr	Thr	Leu	Tyr	Ser 70	Thr	Asn	Leu	Val	Ile 75	Ser	Asp	Ile	Leu	Phe 80
	Thr	Thr	Ala	Leu	Pro 85	Thr	Arg	Ile	Ala	Tyr 90	Tyr	Ala	Met	Gly	Phe 95	Asp
15	Trp	Arg	Ile	Gly 100	Asp	Ala	Leu	Cys	Arg 105	Ile	Thr	Ala	Leu	Val 110	Phe	Tyr
	Ile	Asn	Thr 115	Tyr	Ala	Gly	Val	Asn 120	Phe	Met	Thr	Cys	Leu 125		Ile	Asp
20	Arg	Phe 130	Ile	Ala	Val	Val	His 135	Pro	Leu	Arg	Tyr	Asn 140	Lys	Ile	Lys	Arg
. 1.	Ile 145		His	Ala	Lys	Gly 150	Val	Cys	Ile	Phe	Val 155	Trp	Ile	Leu	Val	Phe 160
	Ala	Gln	Thr	Leu	Pro 165	Leu	Leu	Ile	Asn	Pro 170	Met	Ser	Lys	Gln	Glu 175	Ala
25	Glu	Arg	Ile	Thr 180		Met	Glu	Tyr	Pro 185			Glu	Glu	Thr 190	Lys	Ser
	Leu	Pro	Trp 195	Ile	Leu	Leu	Gly	Ala 200	Cys	Phe	Ile	Gly	Tyr 205	Val	Leu	Pro
30	Leu	Ile 210	Ile	Ile	Leu	Ile	Cys 215	Tyr	Ser	Gln	Ile	Cys 220	Cys	Lys	Leu	Phe
	Arg 225		Ala	Lys	Gln	Asn 230		Leu	Thr	Glu	Lys 235	Ser	Gly	Val	Asn	Lys 240
	Lys	Ala	Lys	Asn	Thr 245		Ile	Leu	Ile	Ile 250		Val	Phe	Val	Leu 255	Cys
35	Phe	Thr	Pro	Tyr 260		Val	Ala	Ile	Ile 265		His	Met	Ile	Lys 270	Lys	Leu

WO 00/22129 PCT/US99/23938

	Arg P	275	ssii Pile	neu Giu	280	GIN Arg	285	Pne Gin	11e
		eu His F 90	he Thr	Val Cys 295	Leu Met	Asn Phe	Asn Cys 300	Cys Met	Asp
5		he Ile T	yr Phe	Phe Ala	Cys Lys	Gly Tyr 315	Lys Arg	Lys Val	Met 320
	Arg M	et Leu L	ys Arg 325	Gln Val	Ser Val	Ser Ile 330	Ser Ser	Ala Val	Lys
10	Ser A		Slu Glu 40	Asn Ser	Arg Glu 345	Met Thr	Glu Thr	Gln Met 350	Met
	Ile H	is Ser L 355	ys Ser	Ser Asn	Gly Lys 360				
	(208) INFO	RMATION	FOR SEC	ID NO:	207:			•	
15		(A) LENG	TH: 144 : nucle	S: singl	pairs				,
	(ii)	MOLECUL	E TYPE:	DNA (ge	enomic)				
20	(xi)	SEQUENC	E DESCR	IPTION:	SEQ ID N	10:207:			
	ATGCGGTGGC	TGTGGCC	CCT GGC	TGTCTCT	CTTGCTGT	GA TTTT	GCTGT G	3GGCTAAGC	60
	AGGGTCTCTG	GGGGTGC	CCC CCI	GCACCTG	GGCAGGCA	CA GAGCO	GAGAC C	CAGGAGCAG	120
	CAGAGCCGAT	CCAAGAG	GGG CAC	CGAGGAT	GAGGAGGC	CA AGGG	CGTGCA G	CAGTATGTG	180
	CCTGAGGAGT	GGGCGGA	GTA CCC	ccggccc	ATTCACCC	TG CTGGC	CCTGCA G	CAACCAAG	240
25	CCCTTGGTGG	CCACCAG	CCC TAA	CCCCGAC	AAGGATGG	GG GCACO	CCAGA C	AGTGGGCAG	300
	GAACTGAGGG	GCAATCT	GAC AGG	GGCACCA	GGGCAGAG	GC TACAG	SATCCA G	AACCCCCTG	360
	TATCCGGTGA	CCGAGAG	CTC CTA	CAGTGCC	TATGCCAT	CA TGCT1	CTGGC G	CTGGTGGTG	420
	TTTGCGGTGG	GCATTGT	GGG CAA	CCTGTCG	GTCATGTG	CA TCGTO	STGGCA C	AGCTACTAC	480
	CTGAAGAGCG	CCTGGAA	CTC CAŢ	CCTTGCC	AGCCTGGC	CC TCTGG	GATTT TO	TGGTCCTC	540
30	TTTTTCTGCC	TCCCTAT	TGT CAT	CTTCAAC	GAGATCAC	CA AGCAG	SAGGCT AG	TGGGTGAC	600
	GTTTCTTGTC	GTGCCGT	GCC CTI	CATGGAG	GTCTCCTC	TC TGGGA	GTCAC G	ACTTTCAG C	660
	CTCTGTGCCC	TGGGCAT	TGA CCG	CTTCCAC	GTGGCCAC	CA GCACO	CTGCC C	\AGGTGAGG	720
	CCCATCGAGC	GGTGCCA	ATC CAT	CCTGGCC	AAGTTGGC	TG TCATO	TGGGT G	GCTCCAT G	780

	ACGCTGGCTG	TGCCTGAGCT	CCTGCTGTGG	CAGCTGGCAC	AGGAGCCTGC	CCCCACCATG	840
	GGCACCCTGG	ACTCATGCAT	CATGAAACCC	TCAGCCAGCC	TGCCCGAGTC	CCTGTATTCA	900
	CTGGTGATGA	CCTACCAGAA	CGCCCGCATG	TGGTGGTACT	TTGGCTGCTA	CTTCTGCCTG	960
	CCCATCCTCT	TCACAGTCAC	CTGCCAGCTG	GTGACATGGC	GGGTGCGAGG	CCCTCCAGGG	1020
5	AGGAAGTCAG	AGTGCAGGGC	CAGCAAGCAC	GAGCAGTGTG	AGAGCCAGCT	CAAGAGCACC	1080
	GTGGTGGGCC	TGACCGTGGT	CTACGCCTTC	TGCACCCTCC	CAGAGAACGT	CTGCAACATC	1140
	GTGGTGGCCT	ACCTCTCCAC	CGAGCTGACC	CGCCAGACCC	TGGACCTCCT	GGGCCTCATC	1200
	AACCAGTTCT	CCACCTTCTT	CAAGGGCGCC	ATCACCCCAG	TGCTGCTCCT	TTGCATCTGC	1260
	AGGCCGCTGG	GCCAGGCCTT	CCTGGACTGC	TGCTGCTGCT	GCTGCTGTGA	GGAGTGCGGC	1320
10	GGGGCTTCGG	AGGCCTCTGC	TGCCAATGGG	TCGGACAACA	AGCTCAAGAC	CGAGGTGTCC	1380
	TCTTCCATCT	ACTTCCACAA	GCCCAGGGAG	TCACCCCCAC	TCCTGCCCCT	GGGCACACCT	1440
	TGCTGA			•			1446

(209) INFORMATION FOR SEQ ID NO:208:

- (i) SEQUENCE CHARACTERISTICS: 15
 - (A) LENGTH: 481 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein
- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:208:

Met Arg Trp Leu Trp Pro Leu Ala Val Ser Leu Ala Val Ile Leu Ala 5

Val Gly Leu Ser Arg Val Ser Gly Gly Ala Pro Leu His Leu Gly Arg

25 . His Arg Ala Glu Thr Gln Glu Gln Gln Ser Arg Ser Lys Arg Gly Thr

Glu Asp Glu Glu Ala Lys Gly Val Gln Gln Tyr Val Pro Glu Glu Trp

Ala Glu Tyr Pro Arg Pro Ile His Pro Ala Gly Leu Gln Pro Thr Lys 30

> Pro Leu Val Ala Thr Ser Pro Asn Pro Asp Lys Asp Gly Gly Thr Pro 90

> Asp Ser Gly Gln Glu Leu Arg Gly Asn Leu Thr Gly Ala Pro Gly Gln

•			÷	100	٠.		•	•	105					110		
	Arg	Leu	Gln 115	Ile	Gln	Asn	Pro	Leu 120	Tyr	Pro	Val	Thr	Glu 125	Ser	Ser	Туз
5	Ser	Ala 130	Tyr	Ala	Ile	Met	Leu 135	Leu	Ala	Leu	Val	Val 140	Phe	Ala	Val	Gl
	Ile 145	Val	Gly	Asn	Leu	Ser 150	Val	Met	Cys	Ile	Val 155	Trp	His	Ser	Tyr	Тут 160
	Leu	Lys	Ser	Ala	Trp 165	Asn	Ser	Ile	Leu	Ala 170	Ser	Leu	Ala	Leu	Trp 175	Asp
10	Phe	Leu	Val	Leu 180	Phe	Phe	Cys	Leu	Pro 185	Ile	Val	Ile	Phe	Asn 190	Glu	Ile
	Thr		Gln 195	Arg	Leu	Leu	Gly	Asp 200	Val	Ser	Cys	Arg	Ala 205	Val	Pro	Phe
15	Met	Glu 210	Val	Ser	Ser	Leu	Gly 215	Val	Thr	Thr	Phe	Ser 220	Leu	Cys	Ala	Leu
	Gly 225	Ile	Asp	Arg	Phe	His 230	Val	Ala	Thr	Ser	Thr 235	Leu	Pro	Lys	Val	Arg 240
	Pro	Ile	Glu	Arg	Cys 245	Gln	Ser	Ile	Leu	Ala 250	Lys	Leu	Ala	Val	Ile 255	Trp
20	Val	Gly	Ser	Met 260	Thr	Leu	Ala	Val	Pro 265		Leu	Leu	Leu	Trp 270	Gln	Leu
	Ala	Gln	Glu 275	Pro	Ala	Pro	Thr	Met 280	Gly	Thr	Leu	Asp	Ser 285	Cys	Ile	Met
25	Lys	Pro 290	Ser	Ala	Ser	Leu	Pro 295	Glu	Ser	Leu	Tyr	Ser 300	Leu	Val	Met	Thr
	Tyr 305	Gln	Asn	Ala	Arg	Met 310	Trp	Trp	Tyr	Phe	Gly 315	Cys	Tyr	Phe	Cys	Leu 320
	Pro	Ile		Phe	Thr 325		Thr	Cys	Gln	Leu 330	Val	Thr	Trp		Val 335	Arg
30	Gly	Pro	Pro	Gly 340	Arg	Lys	Ser	Glu	Cys 345	Arg	Ala	Ser	Lys	His 350	Glu	Gln
	Cys		Ser 355	Gln	Leu	Lys	Ser	Thr 360	Val	Val	Gly	Leu	Thr 365	Val	Val	Tyr
35	Ala	Phe 370	Cys	Thr	Leu	Pro	Glu 375	Asn	Val	Cys	Asn	Ile 380	Val	Val	Ala	Tyr
	Leu 385	Ser		Glu		Thr 390		Gln	Thr	Leu	Asp 395		Leu	Gly	Leu	Ile 400

		Asn	Gln	Phe	Ser	Thr 405	Phe	Phe	Lys	Gly	Ala 410	Ile	Thr	Pro	Val	Leu 415	Leu	
		Leu	Cys	Ile	Cys 420	Arg	Pro	Leu	Gly	Gln 425	Ala	Phe	Leu	Asp	Cys 430	Cys	Cys	
5		Cys	Cys	Cys 435	Cys	Glu	Glu	Cys	Gly 440	Gly	Ala	Ser	Glu	Ala 445	Ser	Ala	Ala	
		Asn	Gly 450	Ser	Asp	Asn	Lys	Leu 455	Lys	Thr	Glu	Val	Ser 460	Ser	Ser	Ile	Tyr	
10		Phe 465	His	Lys	Pro	Arg	Glu 470	Ser	Pro	Pro	Leu	Leu 475	Pro	Leu	Gly	Thr	Pro 480	
		Cys							•									
	(210)	INE	FORM	TION	FOR	SEC) ID	NO: 2	209:									
15		(i)	(A) (B) (C)	LEN TYP STF	CE CH IGTH: PE: n RANDE POLOG	110 ucle DNES)1 ba eic a SS: s	se p cid ingl	pairs	i			·					
	•	(ii	.) MC	LECU	ILE I	YPE:	DNA	(ge	nomi	.c)								
20		(xi	.) SE	QUEN	ICE D	ESCR	IPTI	ON:	SEQ	ID N	10:20	9:			٠			
	ATGTG	GAAC	G CG	acgo	CCAG	CGA	AGAG	CCG	GGGT	'TCAA	CC I	CACA	CTGG	ic co	ACCI	GGAC	!	60
	TGGGA	TGCT	T CC	ccce	GCAA	CGA	CTCG	CTG	GGCG	ACGA	GC I	GCTG	CAGO	T CI	TCCC	CGCG	; 1	.20
	CCGCT	GCTG	G CG	GGCG	TCAC	AGC	CACC	TGC	GTGG	CACT	CT I	CGTG	GTGG	G TA	TCGC	TGGC	: 1	.80
	AACCT	GCTC	A CC	'ATGC	TGGT	GGI	GTCG	CGC	TTCC	GCGA	GC I	GCGC	ACCA	.C CA	CCAA	CCTC	2	40
25	TACCT	GTCC	A GC	ATGG	CCTT	CTC	CGAT	CTG	CTCA	TCTT	CC I	CTGC	ATGC	c cc	TGGA	CCTC	3	00
	GTTCG	CCTC	T GG	CAGI	'ACCG	GCC	CTGG	AAC	TTCG	GCGA	CC I	CCTC	TGCA	A AC	TCTI	'CCAA	. 3	60
	TTCGT	CAGT	'G AG	AGCI	GCAC	CTA	cgcc	ACG	GTGC	TCAC	CA I	CACA	.GCGC	T GA	GCGT	'CGAG	4	20
	CGCTA	CTTC	G CC	ATCT	GCTT	ccc	ACTO	CGG	GCCA	AGGT	'GG 'I	GGTC	ACCA	A GC	GGCG	GGTG	4	80
	AAGCT	GGTC	A TC	TTCG	TCAT	CTG	GGCC	GTG	GCCT	TCTG	CA G	CGCC	GGGC	C CA	TCTT	CGTG	5	40
30	CTAGT	ceee	G TG	GAGC	ACGA	GAA	.CGGC	ACC	GACC	CTTG	GG A	CACC	AACG	A GI	GCCG	cccc	6	00
	ACCGA	GTTT	G CG	GTGC	GCTC	TGG	ACTG	CTC	ACGG	TCAT	GG I	GTGG	GTGT	C CA	GCAT	CTTC	6	60
	TTCTT	CCTT	C CT	GTCT	TCTG	TCT	CACG	GTC	CTCT	ACAG	TC T	CATC	GGCA	G GA	AGCT	GTGG	7	20
	·CGGAG	GAGG	C GC	GGCG	ATGC	TGT	CGTG	GGT	GCCT	CGCT	CA G	GGAC	CAGA	A CC	ACAA	.GCAA	. 7	80

	ACCAA	GAAA	A TG	CTGG	CTGT	AGT	GGTG	TTT	GCCT	TCAT	cc _, 1	CTGC	TGGC	T CC	CCTI	CCAC	8	40
	GTAGG	GCGA	T AT	TATT	TTTC	CAA	ATCC	TTT	GAGC	CTGG	CT C	CTTG	GAGA	T TG	CTCA	GATC	9	00
	AGCCA	GTAC	T GC	AACC	TCGT	GTC	CTTI	GTC	CTCT	TCTA	.cc 1	CAGI	GCTG	C CA	TCAA	cccc	9	60
	ATTCT	GTAC	A AC	ATCA	TGTC	CAA	GAAG	TAC	CGGG	TGGC	AG I	GTTC	AGAC	T TC	TGGG	ATTC	10	20
5	GAACC	CTTC	T CC	CAGA	GAAA	GCT	CTCC	ACT	CTGA	AAGA	TG A	AAGI	TCTC	G GG	CCTG	GACA	. 10	80
	GAATC	TAGT	TT A	'AATA	CATG	A											11	01
	(211)	INF	ORMA	MOIT	FOR	SEQ	ID	NO:2	210:					•				
10			(A) (B) (C) (D)	LEN TYP STR TOP	E CH IGTH: PE: a RANDE POLOG	366 mino DNES Y: n	ami aci S: ot r	no a d	cids vant									
					ILE I								٠		•			;
		•	•	_	ICE I												**	
15		Met 1	Trp	Asn	Ala	Thr 5	Pro	Ser	Glu	Glu	Pro 10	Gly	Phe	Asn	Leu	Thr 15	Leu	•
		Ala	Asp	Leu	Asp 20	Trp	Asp	Ala	Ser	Pro 25	Gly	Asn	Asp	Ser	Leu 30	Gly	Asp	, -
20	. :	Glu	Leu	Leu 35	Gln	Leu	Phe	Pro	Ala 40	Pro	Leu	Leu	Ala	Gly 45	Val	Thr	Ala	
		Thr	Суs 50	Val	Ala	Leu	Phe	Val 55	Val	Gly	Ile	Ala	Gly 60	Asn	Leu	Leu	Thr	
	-	Met 65	Leu	Val	Val	Ser	Arg 70	Phe	Arg	Glu	Leu	Arg 75	Thr	Thr	Thr	Asn	Leu 80	
25		Tyr	Leu	Ser	Ser	Met 85	Ala	Phe	Ser	Asp	Leu 90	Leu	Ile	Phe	Leu	Cys 95	Met	
		Pro	Leu	Asp	Leu 100	Val	Arg	Leu	Trp	Gln 105	Tyr	Arg	Pro	Trp	Asn 110	Phe	Gly	
30		Asp	Leu	Leu 115	_	Lys	Leu	Phe	Gln 120	Phe	Val	Ser	Glu	Ser 125	Cys	Thr	Tyr	
		Ala	Thr 130	Val	Leu	Thr	Ile	Thr 135	Ala	Leu	Ser	Val	Glu 140	Arg	Tyr	Phe	Ala	
		Ile 145	Cys	Phe	Pro	Leu	Arg 150		Lys	Val	Val	Val 155	Thr	Lys	Gly	Arg	Val 160	
35		Lys	Leu	Val	Ile	Phe	Val	Ile	Trp	Ala	Val	Ala	Phe	Cys	Ser	Ala	Gly	

£.

			•			165	٠.		1.74	4,	170		1 ++			175		
		Pro	Ile	Phe	Val 180	Leu	Val	Gly	Val	Glu 185	His	Glu	Asn	Gly	Thr 190	Asp	Pro	
•	5	Trp	Asp	Thr 195	Asn	Glu :	Cys	Arg	Pro 200	Thr	Glu	Phe	Ala	Val 205	Arg	Ser	Gly	
	-	Leu	Leu 210	Thr	Val	Met	Val	Trp 215	Val	Ser	Ser	Ile	Phe 220	Phe	Phe	Leu	Pro	
		Val 225	Phe	Cys	Leu	Thr	Val 230	Leu	Tyr	Ser	Leu	Ile 235	Gly	Arg	Lys		Trp 240	
	10	Arg	Arg	Arg	Arg	Gly 245	Asp	Ala	Val	Val	Gly 250	Ala	Ser	Leu	Arg	Asp 255	Gln	
•	:	Asn	His	Lys	Gln 260	Thr	Lys	Lys	Met	Leu 265	Ala	Val	Val	Val	Phe 270	Ala	Phe	
	15	Ile	Leu	Cys 275	Trp	Leu	Pro	Phe	His 280	Val	Gly	Arg	Tyr	Leu 285	Phe	Ser	Lys	
٠.,		Ser	Phe 290	Glu	Pro	Gly	Ser	Leu 295	Glu	Ile	Ala	Gln	Ile 300	Ser	Gln	Tyr	Cys	
. ,		. Asn 305	Leu	.Val	Ser	Phe	Val 310	Leu	Phe	Tyr	Leu	Ser 315	Ala	Ala	Ile	Asn	Pro 320	
	20	Ile	Leu	Tyr	Asn	Ile 325	Met	Ser	Lys	Lys	Tyr 330	Arg	Val	Ala	Val	Phe 335	Arg	
		Leu	Leu	Gly	Phe 340	Glu	Pro	Phe	Ser	Gln 345	Arg	Lys	Leu	Ser	Thr 350	Leu	Lys	
	25	Asp	Glu ,	Ser 355	Ser	Arg	Ala	Trp		Glu		Ser	Ile	Asn 365				
		(212) IN	FORM	ATIO	N FOI	R SE	Q ID	NO:	211:						•			
		(i) SE	QUEN						s								
	30		(B	TY:	PE: 1	nucl	eic	acid					٠		,			
			(D) TO:	POLO	GY:	line	ar										
		(i	i) M	OLEC	ULE '	TYPE	: DN	A (g	enom	ic)								
		, . (x	i) S	EQUE	NCE I	DESC	RIPT	ION:	SEQ	ID :	NO.: 2	11:	٠٠.					
		ATGCGAGC	CC C	GGGC	GCGC'	T TC	TCGC	CCGC	ATG	TCGC	GGC '	TACT	GC T T	CT G	CTAC	TGCT	2	60
	35	AAGGTGTC	TG C	CTCT	TCTG	c cc	TCGG	GGTC	GCC	CCTG	CGT	CCAG	AAAC	GA A	ACTT	GTCT	3 1	120
		GGGGAGAG	CT G	TGCA	CCTA	C AG	TGAT	CCAG	CGC	CGCG	GCA (GGGA	CGCC	TG G	GGAC	CGGG	A 3	180

	AATTCTGCAA	GAGACGTTCT	GCGAGCCCGA	GCACCCAGGG	AGGAGCAGGG	GGCAGCGTTT	240
	CTTGCGGGAC	CCTCCTGGGA	CCTGCCGGCG	GCCCGGGCC	GTGACCCGGC	TGCAGGCAGA	300
	GGGGCGGAGG	CGTCGGCAGC	CGGACCCCCG	GGACCTCCAA	CCAGGCCACC	TGGCCCCTGG	360
	AGGTGGAAAG	GTGCTCGGGG	TCAGGAGCCT	TCTGAAACTT	TGGGGAGAGG	GAACCCCACG	420
- 5	GCCCTCCAGC	TCTTCCTTCA	GATCTCAGAG	GAGGAAGAGA	AGGGTCCCAG	AGGCGCTGGC	480
	ATTTCCGGGC	GTAGCCAGGA	GCAGAGTGTG	AAGACAGTCC	CCGGAGCCAG	CGATCTTTTT	540
	TACTGGCCAA	GGAGAGCCGG	GAAACTCCAG	GGTTCCCACC	ACAAGCCCCT	GTCCAAGACG	600
	GCCAATGGAC	TGGCGGGGCA	CGAAGGGTGG	ACAATTGCAC	TCCCGGGCCG	GGCGCTGGCC	660
	CAGAATGGAT	CCTTGGGTGA	AGGAATCCAT	GAGCCTGGGG	GTCCCCGCCG	GGGAAACAGC	720
10	ACGAACCGGC	GTGTGAGACT	GAAGAACCCC	TTCTACCCGC	TGACCCAGGA	GTCCTATGGA	780
	GCCTACGCGG	TCATGTGTCT	GTCCGTGGTG	ATCTTCGGGA	CCGGCATCAT	TGGCAACCTG	840
	GCGGTGATGT	GCATCGTGTG	CCACAACTAC	TACATGCGGA	GCATCTCCAA	CTCCCTCTTG	900
	GCCAACCTGG	CCTTCTGGGA	CTTTCTCATC	ATCTTCTTCT	GCCTTCCGCT	GGTCATCTTC	960
	CACGAGCTGA	CCAAGAAGTG	GCTGCTGGAG	GACTTCTCCT	GCAAGATCGT	GCCCTATATA	1020
15	GAGGTCGCCT	CTCTGGGAGT	CACCACTTTC	ACCTTATGTG	CTCTGTGCAT	AGACCGCTTC	1080
	CGTGCTGCCA	CCAACGTACA	GATGTACTAC	GAAATGATCG	AAAATTGTTC	CTCAACAACT	1140
	GCCAAACTTG	CTGTTATATG	GGTGGGAGCT	CTATTGTTAG	CACTTCCAGA	AGTTGTTCTC	1200
	CGCCAGCTGA	GCAAGGAGGA	TTTGGGGTTT	AGTGGCCGAG	CTCCGGCAGA	AAGGTGCATT	1260
	ATTAAGATCT	CTCCTGATTT	ACCAGACACC	ATCTATGTTC	TAGCCCTCAC	CTACGACAGT	1320
20	GCGAGACTGT	GGTGGTATTT	TGGCTGTTAC	TTTTGTTTGC	CCACGCTTTT	CACCATCACC	1380
	TGCTCTCTAG	TGACTGCGAG	GAAAATCCGC	AAAGCAGAGA	AAGCCTGTAC	CCGAGGGAAT	1440
	AAACGGCAGA	TTCAACTAGA	GAGTCAGATG	AAGTGTACAG	TAGTGGCACT	GACCATTTTA	1500
	TATGGATTTT	GCATTATTCC	TGAAAATATC	TGCAACATTG	TTACTGCCTA	CATGGCTACA	1560
	GGGGTTTCAC	AGCAGACAAT	GGACCTCCTT	AATATCATCA	GCCAGTTCCT	TTTGTTCTTT	1620
25	AAGTCCTGTG	TCACCCCAGT	CCTCCTTTTC	TGTCTCTGCA	AACCCTTCAG	TCGGGCCTTC	1680
	ATGGAGTGCT	GCTGCTGTTG	CTGTGAGGAA	TGCATTCAGA	AGTCTTCAAC	GGTGACCAGT	1740
	GATGACAATG	ACAACGAGTA	CACCACGGAA	CTCGAACTCT	CGCCTTTCAG	TACCATACGC	1800
	<i>ርር</i> ተር እ እ እጥርጥ		ምምርምር ምርርርል	ልሮሞሮልሞሞርሮሞ	GD.		1842

	(213) IN	IFORM.	ATIO	N FOI	R SE(QID	NO:2	212:	į.								
5	(i	(B (C	QUENC) LEI) TY!) ST!) TO!	NGTH PE: 8 RANDI	: 613 amino EDNES	3 am: cac: SS:	ino a id	acida	3								
	(i	.i) M	OLEC	ULE :	TYPE:	: pro	oteir	n.									
	(x	i) S	EQUEI	NCE 1	DESCI	RIPT	ON:	SEQ	ID 1	10:23	L2:						
10	, Met	. Arg	Ala	Pro	Gly 5	Ala	Leu	Leu	Ala	Arg 10	Met	Ser	Arg	Leu	Leu 15	Leu	
	Leu	Leu	Leu	Leu 20	Lys	Val	Ser	Ala	Ser 25	Ser	Ala	Leu	Gly	Val 30	Ala	Pro	
	, Ala	Ser	Arg 35	Asn	Glų	Thr	Cys	Leu 40	Gly	Glu	Ser	Cys	Ala 45	Pro	Thr	Val	
15	Ile	Gln 50	Arg	Arg	Gly	Arg	Asp 55	Ala	Trp	Gly	Pro	Gly 60	Asn	Ser	Ala	Arg	
***	Asp 65	Val	Leu	Arg	Ala	Arg 70	Ala	Pro	Arg	Glu	Glu 75	Gln	Gly	Ala	Ala	Phe 80	
20	Leu	Ala	Gly	Pro	Ser 85	Trp	Asp	Leu	Pro	Ala 90	Ala	Pro	Gly	Arg	Asp 95	Pro	
	Ala	Ala	Gly	Arg 100	Gly	Ala	Glu	Ala	Ser 105	Ala	Ala	Gly	Pro	Pro 110	Gly	Pro	
	Pro	Thr	Arg 115	Pro	Pro	Gly	Pro	Trp 120	Arg	Trp	Lys	Gly	Ala 125	Arg	Gly	Gln	
25	Glu	130		Glu	Thr	Leu	Gly 135	Arg	Gly	Asn	Pro	Thr 140	Ala	Leu	Gln	Leu	
	Phe 145	Leu	Gln	Ile	Ser	Glu 150	Glu	Glu	Glu	Lys	Gly 155	Pro	Arg	Gly	Ala	Gly 160	
30	Ile	: Ser	Gly	Arg	Ser 165	Gln	Glu	Gln	Ser	Val 170	Lys	Thr	Val	Pro	Gly 175	Ala	
	Ser	Asp	Leu	Phe 180	Tyr	Trp	Pro	Arg	Arg 185	Ala	Gly	Lys	Leu	Gln 190	Gly	Ser	
	His	His	Lys 195	Pro	Leu	Ser	Lys	Thr 200	Ala	Asn	Gly	Leu	Ala 205	Gly	His	Glu	
35		7 Trp 210					215	-				220			-		
	Lev	Gly	Glu	Gly	Ile	His	Glu	Pro	Gly	Gly	Pro	Arg	Arg	Gly	Asn	Ser	

	225					230	*				235					240
	Thr	Asn	Arg	Arg	Val 245	Arg	Leu	Lys	Asn	Pro 250	Phe	Tyr	Pro	Leu	Thr 255	Gln
5	Glu	Ser	Tyr	Gly 260	Ala	Tyr	Ala	Val	Met 265	Cys	Leu	Ser	Val	Val 270	Ile	Phe
	Gly	Thr	Gly 275	Ile	Ile	Gly	Asn	Leu 280	Ala	Val	Met	Cys	Ile 285	Val	Cys	His
	Asn	Tyr 290	Tyr	Met	Arg	Ser	Ile 295	Ser	Asn	Ser	Leu	Leu 300	Ala	Asn	Leu	Ala
10	Phe 305	Trp	Asp	Phe	Leu	Ile 310	Ile	Phe	Phe	Cys	Leu 315	Pro	Leu	Val	Ile	Phe 320
	His	Glu	Leu	Thr	Lys 325	Lys	Trp	Leu	Leu	Glu 330	Asp	Phe	Ser	Cys	Lys 335	Ile
15	Val	Pro	Tyr	Ile 340	Glu	Val	Ala	Ser	Leu 345	Gly	Val	Thr	Thr	Phe 350	Thr	Leu
	Cys	Ala	Leu 355	Cys	Ile	qzA	Arg	Phe 360	Arg	Ala	Ala	Thr	Asn 365	Val	Gln	Met
	Tyr	Tyr 370	Glu	Met	Ile	Glu	Asn 375	Cys	Ser	Ser	Thr	Thr 380	Ala	Lys	Leu	Ala
20	Val 385	Ile	Trp	Val	Gly	Ala 390	Leu	Leu	Leu	Ala	Leu 395	Pro	Glu	Val	Val	Leu 400
	Arg	Gln	Leu	Ser	Lys 405	Glu	Asp	Leu	Gly	Phe 410	Ser	Gly	Arg	Ala	Pro 415	Ala
25	Glu	Arg	Cys	Ile 420	Ile	Lys	Ile "	Ser	Pro 425	Asp	Leu	Pro	Asp	Thr 430	Ile	Tyr
	Val	Leu	Ala 435	Leu	Thr	Tyr	Asp	Ser 440	Ala	Arg	Leu	Trp	Trp 445	Tyr	Phe	Gly
	Cys	Tyr 450	Phe	Cys	Leu	Pro	Thr 455	Leu	Phe	Thr	Ile	Thr 460	Cys	Ser	Leu	Val
30	Thr 465	Ala	Arg	Lys	Ile	Arg 470	Lys	Ala	Glu	Lys	Ala 475	Cys	Thr	Arg	Gly	Asn 480
	Lys	Arg	Gln	Ile	Gln 485	Leu	Glu	Ser	Gln	Met 490	Lys	Cys	Thr	Val	Val 495	Ala
35	Leu	Thr	Ile	Leu 500	Tyr	Gly	Phe	Cys	Ile 505	Ile	Pro	Glu	Asn	Ile 510	Cys	Asn
	Ile	Val	Thr 515	Ala	туг	Met	Ala	Thr 520	Gly	Val	Ser	Gln	Gln 525	Thr	Met	Asp

		Leu	Leu 530	Asn	Ile	Ile	Ser	Gln 535	Phe	Leu	Leu	Phe	Phe 540	Lys	Ser	Суѕ	Val	
		Thr 545	Pro	Val	Leu	Leu	Phe 550		Leu	Cys	Lys	Pro 555	Phe	Ser	Arg	Ala	Phe 560	
5		Met	Glu.	Cys	Cys	Cys 565	Cys	Cys	Cys	Glu	Glu 570	Cys	Ile	Gln	Lys	Ser 575	Ser	
		Thr	Val	Thr	Ser 580	Asp	Asp	Asn	Asp	Asn 585	Glu	Tyr	Thr	Thr	Glu 590	Leu	Glu	
10		Leu	Ser	Pro 595	Phe	Ser	Thr		Arg 600	Arg	Glu	Met	Ser	Thr 605	Phe	Ala	Ser	
		Val	Gly 610	Thr	His	Cys												
	(214)	INF	FORMA	MOITA	I FOR	SEÇ	D	NO:2	13:									
15		(i)	(B) (C)	LEN TYP STR	CE CH IGTH: PE: n PANDE	124 ucle DNES	8 ba ic a S: s	se p cid singl	airs	;								
		(ii	(ע) MO (.						nomi	c).			*					
20	•		.) SE									.3 :	•					
	ATGGT	TTTT	G CT	CACA	GAAT	GGA	TAAC	AGC	AAGC	CACA	TT 1	GATT	'ATTC	C TA	CACI	TCTG	;	6
	GTGCC	CCTC	C AA	AACC	GCAG	CTG	CACT	'GAA	ACAG	CCAC	AC C	TCTG	CCAA	.G CC	ATA.	CCTG	; 1	.2
	ATGGA	AT _. TA	A GT	'GAĠG	AGCA	. CAG	TTGG	ATG	AGCA	ACCA	AA C	AGAC	CTTC	A CI	ATGI	GCTG	, 1	.8
	AAACC	CGGG	G AA	.GTGG	CCAC	AGC	CAGC	ATC	TTCT	TTGG	GA I	TCTG	TGGT	T GI	TTTC	TATO	2	4
25	TTCGG	CAAT	T CC	CTGG	TTTG	TTT	GGTC	ATC	CATA	.GGAG	TA G	GAGG	ACTC	A GI	CTAC	CACC	: 3	0
	AACTA	CTTT.	G TG	GTCT	CCAT	GGC	ATGT	GCT	GACC	TTCT	CA I	CAGC	GTTG	C CA	GCAC	GCCT	' з	6
	TTCGT	CCTG	C TC	CAGT	TCAC	CAC	TGGA	AGG	TGGA	CGCT	GG G	TAGT	GCAA	C GI	GCAA	GGTT.	4	2
	GTGCG	ATAT	T TT	CAAT	ATCT	CAC	TCCA	.GGT	GTCC	AGAT	CT A	CGTT	CTCC	т ст	CCAT	CTGC	4	8
	ATAGA	.CCGG	T TC	TACA	.CCAT	CGT	CTAT	CCT	CTGA	GCTT	CA A	.GGTG	TCCA	G AG	AAAA	AGCC	5	4
30	AAGAA	AATG	A TT	GCGG	CATC	GTG	GATC	TTT	GATG	CAGG	CT I	TGTG	ACCC	C TG	TGCT	CTTT	6	0
	TTCTA	TGGC	T CC	AACT	GGGA	CAG	TCAT	TGT	AACT.	ATTT	CC T	cccc	тсст	C TT	GGGA	AGGC	6	6
	ACTGC	CTAC	A CT	GTCA	TCCA	CTT	CTTG	GTG	GGCT	TTGT	GA T	TCCA	TCTG	T CC	TCAT	AATT	7	2
	TTATT	TTAC	C AA	AAGG	TCAT	AAA	ATAT	ATT	TGGA	GAAT	AG G	רמרמי	באדכ	G CC	CDDC	ሮርጥር	7	٥

WO 00/22129 PCT/US99/23938

	AGGAGGACAA TGAACATTGT CCCTCGGACA AAAGTGAAAA CTAAAAAGAT GTTCCTCATT 8	40
	TTAAATCTGT TGTTTTTGCT CTCCTGGCTG CCTTTTCATG TAGCTCAGCT ATGGCACCCC	00
	CATGAACAAG ACTATAAGAA AAGTTCCCTT GTTTTCACAG CTATCACATG GATATCCTTT	60
	AGTTCTTCAG CCTCTAAACC TACTCTGTAT TCAATTTATA ATGCCAATTT TCGGAGAGGG 10	20
5	ATGAAAGAGA CTTTTTGCAT GTCCTCTATG AAATGTTACC GAAGCAATGC CTATACTATC 10	80
	ACAACAAGTT CAAGGATGGC CAAAAAAAAC TACGTTGGCA TTTCAGAAAT CCCTTCCATG 11	40
	GCCAAAACTA TTACCAAAGA CTCGATCTAT GACTCATTTG ACAGAGAAGC CAAGGAAAAA 12	00
	AAGCTTGCTT GGCCCATTAA CTCAAATCCA CCAAATACTT TTGTCTAA 12	48
	(215) INFORMATION FOR SEQ ID NO:214:	
10	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 415 amino acids(B) TYPE: amino acid	
	<pre>(C) STRANDEDNESS: (D) TOPOLOGY: not relevant</pre>	
15	(ii) MOLECULE TYPE: protein	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:214:	
	Met Val Phe Ala His Arg Met Asp Asn Ser Lys Pro His Leu Ile Ile 1 5 10 15	
20	Pro Thr Leu Leu Val Pro Leu Gln Asn Arg Ser Cys Thr Glu Thr Ala 20 25 30	
	Thr Pro Leu Pro Ser Gln Tyr Leu Met Glu Leu Ser Glu Glu His Ser 35 40 45	
	Trp Met Ser Asn Gln Thr Asp Leu His Tyr Val Leu Lys Pro Gly Glu 50 55 60	
25	Val Ala Thr Ala Ser Ile Phe Phe Gly Ile Leu Trp Leu Phe Ser Ile 65 70 75 80	
	Phe Gly Asn Ser Leu Val Cys Leu Val Ile His Arg Ser Arg Arg Thr 85 90 95	
30	Gln Ser Thr Thr Asn Tyr Phe Val Val Ser Met Ala Cys Ala Asp Leu 100 105 110	
•	Leu Ile Ser Val Ala Ser Thr Pro Phe Val Leu Leu Gln Phe Thr Thr 115 120 125	
	Gly Arg Trp Thr Leu Gly Ser Ala Thr Cys Lys Val Val Arg Tyr Phe 130 135 140	

Val Ile His Phe Leu Val Gly Phe Val Ile Pro Ser Val Leu Ile Ile 225 Leu Phe Tyr Gln Lys Val Ile Lys Tyr Ile Trp Arg Ile Gly Thr Asp 245 Gly Arg Thr Val Arg Arg Thr Met Asn Ile Val Pro Arg Thr Lys Val 260 Lys Thr Lys Lys Met Phe Leu Ile Leu Asn Leu Leu Phe Leu Leu Ser 285 Trp Leu Pro Phe His Val Ala Gln Leu Trp His Pro His Glu Gln Asp 290 Tyr Lys Lys Ser Ser Leu Val Phe Thr Ala Ile Thr Trp Ile Ser Phe 305 Ser Ser Ser Ala Ser Lys Pro Thr Leu Tyr Ser Ile Tyr Asn Ala Asn 325 Phe Arg Arg Gly Met Lys Glu Thr Phe Cys Met Ser Ser Met Lys Cys 340 Tyr Arg Ser Asn Ala Tyr Thr Ile Thr Thr Ser Ser Arg Met Ala Lys Thr Ile Lys Asn Tyr Val Gly Ile Ser Glu Ile Pro Ser Met Ala Lys Thr Ile		Gln 145	Tyr	Leu	Thr	Pro	Gly 150	Val	Gln	·Ile	Tyr	Val 155	Leu	Leu	Ser	Ile	Cys 160
Gly Phe Val Thr Pro Val Leu Phe Phe Tyr Gly Ser Asn Trp Asp Ser 205 His Cys Asn Tyr Phe Leu Pro Ser Ser Trp Glu Gly Thr Ala Tyr Thr 210 Val Ile His Phe Leu Val Gly Phe Val Ile Pro Ser Val Leu Ile Ile 225 Leu Phe Tyr Gln Lys Val Ile Lys Tyr Ile Trp Arg Ile Gly Thr Asp 245 Gly Arg Thr Val Arg Arg Thr Met Asn Ile Val Pro Arg Thr Lys Val 260 Lys Thr Lys Lys Met Phe Leu Ile Leu Asn Leu Leu Phe Leu Leu Ser 280 Trp Leu Pro Phe His Val Ala Gln Leu Trp His Pro His Glu Gln Asp 295 Tyr Lys Lys Ser Ser Leu Val Phe Thr Ala Ile Thr Trp Ile Ser Phe 305 Ser Ser Ser Ala Ser Lys Pro Thr Leu Tyr Ser Ile Tyr Asn Ala Asn 325 Phe Arg Arg Gly Met Lys Glu Thr Phe Cys Met Ser Ser Met Ala Lys 355 Tyr Arg Ser Asn Ala Tyr Thr Ile Thr Thr Ser Ser Arg Met Ala Lys 355 Lys Asn Tyr Val Gly Ile Ser Glu Ile Pro Ser Met Ala Lys Thr Ile 370 Thr Lys Asp Ser Ile Tyr Asp Ser Phe Asp Arg Glu Ala Lys Glu Lys 385 Thr Lys Asp Ser Ile Tyr Asp Ser Phe Asp Arg Glu Ala Lys Glu Lys 385 Lys Leu Ala Trp Pro Ile Asn Ser Asn Pro Pro Asn Thr Phe Val		Ile	Asp	Arg	Phe		Thr	Ile	Val	Tyr		Leu	Ser	Phe	Lys		Ser
His Cys Asn Tyr Phe Leu Pro Ser Ser Trp Glu Gly Thr Ala Tyr Thr 215 Val Ile His Phe Leu Val Gly Phe Val Ile Pro Ser Val Leu Ile Ile 225 Leu Phe Tyr Gln Lys Val Ile Lys Tyr Ile Trp Arg Ile Gly Thr Asp 255 Gly Arg Thr Val Arg Arg Thr Met Asn Ile Val Pro Arg Thr Lys Val 266 Lys Thr Lys Lys Met Phe Leu Ile Leu Asn Leu Leu Phe Leu Leu Ser 275 Trp Leu Pro Phe His Val Ala Gln Leu Trp His Pro His Glu Gln Asp 295 Try Lys Lys Ser Ser Leu Val Phe Thr Ala Ile Thr Trp Ile Ser Phe 305 Ser Ser Ser Ala Ser Lys Pro Thr Leu Tyr Ser Ile Tyr Asn Ala Asn 325 Tyr Arg Ser Asn Ala Tyr Thr Ile Thr Thr Ser Ser Arg Met Ala Lys 355 Thr Lys Asp Ser Ile Tyr Asp Ser Phe Asp Arg Glu Ala Lys Glu Lys 385 Thr Lys Asp Ser Ile Tyr Asp Ser Phe Asp Arg Glu Ala Lys Glu Lys 385 Lys Leu Ala Trp Pro Ile Asn Ser Asn Pro Pro Asn Thr Phe Val	5	Arg	Glu	Lys		Lys	Lys	Met	Ile		Ala	Ser	Trp	Ile		Asp	Ala
10 210 215 220 Val Ile His Phe Leu Val Gly Phe Val Ile Pro Ser Val Leu Ile Ile 225 230 Leu Phe Tyr Gln Lys Val Ile Lys Tyr Ile Trp Arg Ile Gly Thr Asp 245 Gly Arg Thr Val Arg Arg Thr Met Asn Ile Val Pro Arg Thr Lys Val 260 Lys Thr Lys Lys Met Phe Leu Ile Leu Asn Leu Leu Phe Leu Leu Ser 285 Trp Leu Pro Phe His Val Ala Gln Leu Trp His Pro His Glu Gln Asp 290 Tyr Lys Lys Ser Ser Leu Val Phe Thr Ala Ile Thr Trp Ile Ser Phe 305 Ser Ser Ser Ala Ser Lys Pro Thr Leu Tyr Ser Ile Tyr Asn Ala Asn 325 Phe Arg Arg Gly Met Lys Glu Thr Phe Cys Met Ser Ser Met Lys Cys 340 Tyr Arg Ser Asn Ala Tyr Thr Ile Thr Thr Ser Ser Arg Met Ala Lys 355 Lys Asn Tyr Val Gly Ile Ser Glu Ile Pro Ser Met Ala Lys Thr Ile 370 Thr Lys Asp Ser Ile Tyr Asp Ser Phe Asp Arg Glu Ala Lys Glu Lys 385 Lys Leu Ala Trp Pro Ile Asn Ser Asn Pro Pro Asn Thr Phe Val		Gly	Phe		Thr	Pro	Val	Leu		Phe	Tyr	Gly	Ser		Trp	Asp	Ser
Leu Phe Tyr Gln Lys Val Ile Lys Tyr Ile Trp Arg Ile Gly Thr Asp 245 Gly Arg Thr Val Arg Arg Thr Met Asn Ile Val Pro Arg Thr Lys Val 260 Lys Thr Lys Lys Met Phe Leu Ile Leu Asn Leu Leu Phe Leu Leu Ser 275 Trp Leu Pro Phe His Val Ala Gln Leu Trp His Pro His Glu Gln Asp 290 Tyr Lys Lys Ser Ser Leu Val Phe Thr Ala Ile Thr Trp Ile Ser Phe 305 Ser Ser Ser Ala Ser Lys Pro Thr Leu Tyr Ser Ile Tyr Asn Ala Asn 325 Phe Arg Arg Gly Met Lys Glu Thr Phe Cys Met Ser Ser Met Lys Cys 345 Tyr Arg Ser Asn Ala Tyr Thr Ile Thr Thr Ser Ser Arg Met Ala Lys 355 Lys Asn Tyr Val Gly Ile Ser Glu Ile Pro Ser Met Ala Lys Thr Ile 370 Thr Lys Asp Ser Ile Tyr Asp Ser Phe Asp Arg Glu Ala Lys Glu Lys 385 Lys Leu Ala Trp Pro Ile Asn Ser Asn Pro Pro Asn Thr Phe Val	10	His		Asn	Tyr	Phe	Leu		Ser	Ser	Trp	Glu	_	Thr	Ala	Tyr	Thr
245 250 255 Gly Arg Thr Val Arg Arg Thr Met Asn Ile Val Pro Arg Thr Lys Val 260 Lys Thr Lys Lys Met Phe Leu Ile Leu Asn Leu Leu Phe Leu Leu Ser 285 Trp Leu Pro Phe His Val Ala Gln Leu Trp His Pro His Glu Gln Asp 290 Tyr Lys Lys Ser Ser Leu Val Phe Thr Ala Ile Thr Trp Ile Ser Phe 305 Ser Ser Ser Ala Ser Lys Pro Thr Leu Tyr Ser Ile Tyr Asn Ala Asn 325 Phe Arg Arg Gly Met Lys Glu Thr Phe Cys Met Ser Ser Met Lys Cys 340 Tyr Arg Ser Asn Ala Tyr Thr Ile Thr Thr Ser Ser Arg Met Ala Lys 355 Lys Asn Tyr Val Gly Ile Ser Glu Ile Pro Ser Met Ala Lys Thr Ile 370 Thr Lys Asp Ser Ile Tyr Asp Ser Phe Asp Arg Glu Ala Lys Glu Lys 385 Lys Leu Ala Trp Pro Ile Asn Ser Asn Pro Pro Asn Thr Phe Val			Ile	His	Phe	Leu		Gly	Phe	Val	Ile		Ser	Val	Leu	Ile	
Lys Thr Lys Lys Met Phe Leu Ile Leu Asn Leu Leu Phe Leu Leu Ser 285 Trp Leu Pro Phe His Val Ala Gln Leu Trp His Pro His Glu Gln Asp 295 Tyr Lys Lys Ser Ser Leu Val Phe Thr Ala Ile Thr Trp Ile Ser Phe 305 Ser Ser Ser Ala Ser Lys Pro Thr Leu Tyr Ser Ile Tyr Asn Ala Asn 325 Phe Arg Arg Gly Met Lys Glu Thr Phe Cys Met Ser Ser Met Lys Cys 340 Tyr Arg Ser Asn Ala Tyr Thr Ile Thr Thr Ser Ser Arg Met Ala Lys 355 Lys Asn Tyr Val Gly Ile Ser Glu Ile Pro Ser Met Ala Lys Thr Ile 370 Thr Lys Asp Ser Ile Tyr Asp Ser Phe Asp Arg Glu Ala Lys Glu Lys 395 Lys Leu Ala Trp Pro Ile Asn Ser Asn Pro Pro Asn Thr Phe Val	2 · •	Leu	Phe	Tyr	Gln		Val	Ile	Lys	Tyr		Trp	Arg	Ile	Gly		Asp
275 280 285 285 Trp Leu Pro Phe His Val Ala Gln Leu Trp His Pro His Glu Gln Asp 295 Tyr Lys Lys Ser Ser Leu Val Phe Thr Ala Ile Thr Trp Ile Ser Phe 305 Ser Ser Ser Ala Ser Lys Pro Thr Leu Tyr Ser Ile Tyr Asn Ala Asn 335 Phe Arg Gly Met Lys Glu Thr Phe Cys Met Ser Ser Met Lys Cys 345 Tyr Arg Ser Asn Ala Tyr Thr Ile Thr Thr Ser Ser Arg Met Ala Lys 355 Lys Asn Tyr Val Gly Ile Ser Glu Ile Pro Ser Met Ala Lys Thr Ile 370 Thr Lys Asp Ser Ile Tyr Asp Ser Phe Asp Arg Glu Ala Lys Glu Lys 385 Lys Leu Ala Trp Pro Ile Asn Ser Asn Pro Pro Asn Thr Phe Val	15	Gly	Arg	Thr		Arg	Arg	Thr	Met		Ile	Val	Pro	Arg		Lys	Val
290 295 300 Tyr Lys Lys Ser Ser Leu Val Phe Thr Ala Ile Thr Trp Ile Ser Phe 305 Ser Ser Ser Ala Ser Lys Pro Thr Leu Tyr Ser Ile Tyr Asn Ala Asn 325 Phe Arg Arg Gly Met Lys Glu Thr Phe Cys Met Ser Ser Met Lys Cys 340 Tyr Arg Ser Asn Ala Tyr Thr Ile Thr Thr Ser Ser Arg Met Ala Lys 355 Lys Asn Tyr Val Gly Ile Ser Glu Ile Pro Ser Met Ala Lys Thr Ile 370 Thr Lys Asp Ser Ile Tyr Asp Ser Phe Asp Arg Glu Ala Lys Glu Lys 385 Lys Leu Ala Trp Pro Ile Asn Ser Asn Pro Pro Asn Thr Phe Val				275					280					285	,		
305 310 315 320 Ser Ser Ser Ala Ser Lys Pro Thr Leu Tyr Ser Ile Tyr Asn Ala Asn 325 Phe Arg Arg Gly Met Lys Glu Thr Phe Cys Met Ser Ser Met Lys Cys 340 Tyr Arg Ser Asn Ala Tyr Thr Ile Thr Thr Ser Ser Arg Met Ala Lys 355 Lys Asn Tyr Val Gly Ile Ser Glu Ile Pro Ser Met Ala Lys Thr Ile 375 Thr Lys Asp Ser Ile Tyr Asp Ser Phe Asp Arg Glu Ala Lys Glu Lys 385 Lys Leu Ala Trp Pro Ile Asn Ser Asn Pro Pro Asn Thr Phe Val	20		290					295					300				
Phe Arg Arg Gly Met Lys Glu Thr Phe Cys Met Ser Ser Met Lys Cys 340 Tyr Arg Ser Asn Ala Tyr Thr Ile Thr Thr Ser Ser Arg Met Ala Lys 355 Lys Asn Tyr Val Gly Ile Ser Glu Ile Pro Ser Met Ala Lys Thr Ile 370 Thr Lys Asp Ser Ile Tyr Asp Ser Phe Asp Arg Glu Ala Lys Glu Lys 385 Lys Leu Ala Trp Pro Ile Asn Ser Asn Pro Pro Asn Thr Phe Val		305					310					315					320
Tyr Arg Ser Asn Ala Tyr Thr Ile Thr Thr Ser Ser Arg Met Ala Lys 355 Lys Asn Tyr Val Gly Ile Ser Glu Ile Pro Ser Met Ala Lys Thr Ile 375 Thr Lys Asp Ser Ile Tyr Asp Ser Phe Asp Arg Glu Ala Lys Glu Lys 385 Lys Leu Ala Trp Pro Ile Asn Ser Asn Pro Pro Asn Thr Phe Val						325					330				•	335	
Lys Asn Tyr Val Gly Ile Ser Glu Ile Pro Ser Met Ala Lys Thr Ile 375 Thr Lys Asp Ser Ile Tyr Asp Ser Phe Asp Arg Glu Ala Lys Glu Lys 385 Lys Leu Ala Trp Pro Ile Asn Ser Asn Pro Pro Asn Thr Phe Val	25				340					345				,	350		
Thr Lys Asp Ser Ile Tyr Asp Ser Phe Asp Arg Glu Ala Lys Glu Lys 385 390 395 400 Lys Leu Ala Trp Pro Ile Asn Ser Asn Pro Pro Asn Thr Phe Val				355					360					365			
385 390 395 400 Lys Leu Ala Trp Pro Ile Asn Ser Asn Pro Pro Asn Thr Phe Val	30	-	370					375					380			ė	
4.05		385		ij.			390		٠			395		٠			
	25	Lys	Leu	Ala	Trp		Ile	Asn	Ser	Asn		Pro	Asn	Thr	Phe		

35 (216) INFORMATION FOR SEQ ID NO:215:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1842 base pairs

PCT/US99/23938

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- 5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:215:

ATGGGGCCCA CCCTAGCGGT TCCCACCCC TATGGCTGTA TTGGCTGTAA GCTACCCCAG 60 CCAGAATACC CACCGGCTCT AATCATCTTT ATGTTCTGCG CGATGGTTAT CACCATCGTT 120 GTAGACCTAA TCGGCAACTC CATGGTCATT TTGGCTGTGA CGAAGAACAA GAAGCTCCGG 180 AATTCTGGCA ACATCTTCGT GGTCAGTCTC TCTGTGGCCG ATATGCTGGT GGCCATCTAC 240 CCATACCCTT TGATGCTGCA TGCCATGTCC ATTGGGGGGCT GGGATCTGAG CCAGTTACAG 300 TGCCAGATGG TCGGGTTCAT CACAGGGCTG AGTGTGGTCG GCTCCATCTT CAACATCGTG 360 GCAATCGCTA TCAACCGTTA CTGCTACATC TGCCACAGCC TCCAGTACGA ACGGATCTTC 420 AGTGTGCGCA ATACCTGCAT CTACCTGGTC ATCACCTGGA TCATGACCGT CCTGGCTGTC 480 CTGCCCAACA TGTACATTGG CACCATCGAG TACGATCCTC GCACCTACAC CTGCATCTTC 540⁻ AACTATCTGA ACAACCCTGT CTTCACTGTT ACCATCGTCT GCATCCACTT CGTCCTCCCT 600 CTCCTCATCG TGGGTTTCTG CTACGTGAGG ATCTGGACCA AAGTGCTGGC GGCCCGTGAC 660 CCTGCAGGGC AGAATCCTGA CAACCAACTT GCTGAGGTTC GCAATAAACT AACCATGTTT 720 GTGATCTTCC TCCTCTTTGC AGTGTGCTGG TGCCCTATCA ACGTGCTCAC TGTCTTGGTG 780 GCTGTCAGTC CGAAGGAGAT GGCAGGCAAG ATCCCCAACT GGCTTTATCT TGCAGCCTAC 840 20 TTCATAGCCT ACTTCAACAG CTGCCTCAAC GCTGTGATCT ACGGGCTCCT CAATGAGAAT 900 TTCCGAAGAG AATACTGGAC CATCTTCCAT GCTATGCGGC ACCCTATCAT ATTCTTCTCT 960 GGCCTCATCA GTGATATTCG TGAGATGCAG GAGGCCCGTA CCCTGGCCCG CGCCCGTGCC 1020 CATGCTCGCG ACCAAGCTCG TGAACAAGAC CGTGCCCATG CCTGTCCTGC TGTGGAGGAA 1080 ACCCCGATGA ATGTCCGGAA TGTTCCATTA CCTGGTGATG CTGCAGCTGG CCACCCCGAC 1140 25 CGTGCCTCTG GCCACCCTAA GCCCCATTCC AGATCCTCCT CTGCCTATCG CAAATCTGCC 1200 TCTACCCACC ACAAGTCTGT CTTTAGCCAC TCCAAGGCTG CCTCTGGTCA CCTCAAGCCT 1260 GTCTCTGGCC ACTCCAAGCC TGCCTCTGGT CACCCCAAGT CTGCCACTGT CTACCCTAAG 1320 CCTGCCTCTG TCCATTTCAA GGCTGACTCT GTCCATTTCA AGGGTGACTC TGTCCATTTC 1380 AAGCCTGACT CTGTTCATTT CAAGCCTGCT TCCAGCAACC CCAAGCCCAT CACTGGCCAC 1440

	CATGTCTCTG CTGGCAGCCA CTCCAAGTCT GCCTTCAATG CTGCCACCAG CCACCCTAAA 1	500
	CCCATCAAGC CAGCTACCAG CCATGCTGAG CCCACCACTG CTGACTATCC CAAGCCTGCC 1	560
	ACTACCAGCC ACCCTAAGCC CGCTGCTGCT GACAACCCTG AGCTCTCTGC CTCCCATTGC 1	€20
	CCCGAGATCC CTGCCATTGC CCACCCTGTG TCTGACGACA GTGACCTCCC TGAGTCGGCC 1	680
5	TCTAGCCCTG CCGCTGGGCC CACCAAGCCT GCTGCCAGCC AGCTGGAGTC TGACACCATC 1	740
	GCTGACCTTC CTGACCCTAC TGTAGTCACT ACCAGTACCA ATGATTACCA TGATGTCGTG 1	800
	GTTGTTGATG TTGAAGATGA TCCTGATGAA ATGGCTGTGT GA	842
	(217) INFORMATION FOR SEQ ID NO:216:	
10	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 613 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: (D) TOPOLOGY: not relevant	
	(ii) MOLECULE TYPE: protein	
15	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:216:	
	Met Gly Pro Thr Leu Ala Val Pro Thr Pro Tyr Gly Cys Ile Gly Cys 1 5 10 15	
	Lys Leu Pro Gln Pro Glu Tyr Pro Pro Ala Leu Ile Ile Phe Met Phe 20 25 30	
20	Cys Ala Met Val Ile Thr Ile Val Val Asp Leu Ile Gly Asn Ser Met 35 40 45	
	Val Ile Leu Ala Val Thr Lys Asn Lys Lys Leu Arg Asn Ser Gly Asn 50 55 60	
25	Ile Phe Val Val Ser Leu Ser Val Ala Asp Met Leu Val Ala Ile Tyr 65 70 75 80	
	Pro Tyr Pro Leu Met Leu His Ala Met Ser Ile Gly Gly Trp Asp Leu 85 90 95	
	Ser Gln Leu Gln Cys Gln Met Val Gly Phe Ile Thr Gly Leu Ser Val	
30	Val Gly Ser Ile Phe Asn Ile Val Ala Ile Ala Ile Asn Arg Tyr Cys 115 120 125	
	Tyr Ile Cys His Ser Leu Gln Tyr Glu Arg Ile Phe Ser Val Arg Asn 130 135 140	
35	Thr Cys Ile Tyr Leu Val Ile Thr Trp Ile Met Thr Val Leu Ala Val 145 150 155 160	

		Leu	Pro	Asn	Met 	Tyr 165	Ile	Gly	Thr	Ile	Glu 170	Tyr	Asp	Pro	Arģ	Thr 175	Tyr
		Thr	Cys	Ile	Phe 180	Asn	Tyr	Leu	Asn	Asn 185	Pro	Val	Phe	Thr	Val 190	Thr	Ile
5		Val	Cys	Ile 195	His	Phe	Val	Leu	Pro 200	Leu	Leu	Ile	Val	Gly 205	Phe	Cys	Tyr
	** *	Val	Arg 210	Ile :	Trp	Thr	Lys	Val 215	Leu	Ala	Ala	Arg	Asp 220	Pro	Ala	Gly	Gln
10		Asn 225	Pro	Asp	Asn	Gln	Leu 230	Ala	Glu	Val	Arg	Asn 235	Lys	Leu	Thr	Met	Phe 240
		Val	Ile	Phe	Leu	Leu 245	Phe	Ala	Val	Cys	Trp 250	Cys	Pro	Ile	Asn	Val 255	Leu
		Thr	Val	Leu	Val 260	Ala	Val	Ser	Pro	Lys 265	Glu	Met	Ala	Gly	Lys 270	Ile	Pro
15		Asn	Trp	Leu 275	Tyr	Leu	Ala	Ala	Tyr 280	Phe	Ile	Ala	Tyr	Phe 285		Ser	Cys
4.5		Leu	Asn 290	Ala	Val	Ile	Tyr	Gly 295	Leu	Leu	Asn	Glu	Asn 300	Phe	Arg	Arg	Glu
20		Tyr 305	Trp	Thr	Ile	Phe	His 310	Ala	Met	Arg	His	Pro 315	Ile	Ile	Phe	Phe	Ser 320
		Gly	Leu	Ile	Ser	Asp 325	Ile	Arg	Glu	Met	Gln 330	Glu	Ala	Arg	Thr	Leu 335	Ala
÷		Arg	Ala	Arg	Ala 340	His	Ala	Arg	Asp	Gln 345	Ala	Arg	Glu	Gln	Asp 350	Arg	Ala
25		His	Ala	Cys 355	Pro	Ala	Val	Glu	Glu 360	Thr	Pro	Met	Asn	Val 365	Arg	Asn	Val
		Pro	Leu 370	Pro	Gly	Asp	Ala	Ala 375	Ala	Gly	His	Pro	Asp 380	Arg	Ala	Ser	Gly
30		His 385	Pro	Lys	Pro	His	Ser 390	Arg	Ser	Ser	Ser	Ala 395	Tyr	Arg	Lys	Ser	Ala 400
		Ser	Thr	His	His	Lys 405	Ser	Val	Phe	Ser	His 410	Ser	Lys	Ala	Ala	Ser 415	Gly
		His	Leu	Lys	Pro 420	Val	Ser	Gly	His	Ser 425	Lys	Pro	Ala	Ser	Gly 430	His	Pro
35		Lys	Ser	Ala 435	Thṛ	Val	Tyr	Pro	Lys 440	Pro	Ala	Ser	Val	His 445	Phe	Lys	Ala
		Asp	Ser	Val	His	Phe	Lys	Gly	Asp	Ser	Val	His	Phe	Lys	Pro	Asp	Ser

		450		•			455		٠,			460				
	Val 465	His	Phe	Lys	Pro	Ala 470	Ser	Ser	Asn	Pro	Lys 475	Pro	Ile	Thr	Gly	His 480
5	His	Val	Ser	Ala	Gly 485	Ser	His	Ser	Lys	Ser 490	Ala	Phe	Asn	Ala	Ala 495	Thr
	Ser	His	Pro	Lys 500	Pro	Ile	Lys	Pro	Ala 505	Thr	Ser	His	Ala	Glu 510	Pro	Thr
	Thr	Ala	Asp 515	Tyr	Pro	Lys	Pro	Ala 520	Thr	Thr	Ser	His	Pro 525	Lys	Pro	Ala
10	Ala	Ala 530	Asp	Asn	Pro	Glu	Leu 535	Ser	Ala	Ser	His	Cys 540	Pro	Glu	Ile	Pro
•	Ala 545	Ile	Ala	His	Pro	Val 550	Ser	Asp	Asp	Ser	Asp 555	Leu	Pro	Glu	Ser	Ala 560
15	Ser	Ser	Pro	Ala	Ala 565	Gly	Pro	Thr	Lys	Pro 570	Ala	Ala	Ser	Gln	Leu 575	Glu
	Ser	Asp	Thr	Ile 580	Ala	Asp	Leu	Pro	Asp 585	Pro	Thr	Val	Val	Thr 590	Thr	Ser
	Thr	Asn	Asp 595	Tyr	His	Asp	Val	Val 600	Val	Val	Asp	Val	Glu 605	Asp	Asp	Pro
20	Asp	Glu 610	Met	Ala	Val											
	(218) IN	FORM	MOITA	FOF	SEÇ] ID	NO:2	17:								
25	(i	(B) (C)	LENC TYP STR TOP	IGTH: PE: n LANDE	185 ucle DNES	4 ba ic a S: s	se p cid ingl	airs							·	
	(i:	i) MC	LECU	LE I	YPE:	DNA	(ge	nomi	c)							
	(x:	i) SE	QUEN	CE D	ESCR	IPTI	ON:	SEQ	ID N	0:21	7:					
30	ATGGGGCCC	CA CC	CTAG	CGGT	TCC	CACC	ccc	TATG	GCTG	TA T	TGGC	TGTA	A GC	TACC	CCAG	6
	CCAGAATA	CC CA	CCGG	CTCT	AAT	CATC	TTT	atgt	TCTG	CG C	GATG	GTTA	T CA	CCAT	CGTT	12
	GTAGACCT	AA TO	:GGCA	ACTC	CAT	GGTC	ATT	TTGG	CTGT	GA C	GAAG.	AACA	A GA	AGCT	CCGG	18
	AATTCTGGG	CA AC	ATCT	TCGT	GGT	CAGT	CTC	TCTG	TGGC	CG A	TATG	CTGG	T GG	CCAT	CTAC	24
	CCATACCCT	TT TG	ATGC	TGCA	TGC	CATG	TCC	ATTG	GGGG	CT G	GGAT	CTGA	G CC	AGTT	ACAG	30
35	TGCCAGATO	G TC	GGGT	TCAT	CAC	AGGG	CTG .	AGTG	TGGT	CG G	CTCC	ATCT	T CA	ACAT	CGTG	36

WO 00/22129 PCT/US99/23938

179

	GCAATCGCTA	TCAACCGTTA	CIGCIACAIC	TGCCACAGCC	TCCAGTACGA	ACGGATCTTC	420
	AGTGTGCGCA	ATACCTGCAT	CTACCTGGTC	ATCACCTGGA	TCATGACCGT	CCTGGCTGTC	480
	CTGCCCAACA	TGTACATTGG	CACCATCGAG	TACGATCCTC	GCACCTACAC	CTGCATCTTC	540
	AACTATCTGA	ACAACCCTGT	CTTCACTGTT	ACCATCGTCT	GCATCCACTT	CGTCCTCCCT	.600
5	CTCCTCATCG	TGGGTTTCTG	CTACGTGAGG	ATCTGGACCA	AAGTGCTGGC	GGCCCGTGAC	660
	CCTGCAGGGC	AGAATCCTGA	CAACCAACTT	GCTGAGGTTC	GCAATAAACT	AACCATGTTT	720
	GTGATCTTCC	TCCTCTTTGC	AGTGTGCTGG	TGCCCTATCA	ACGTGCTCAC	TGTCTTGGTG	780
	GCTGTCAGTC	CGAAGGAGAT	GGCAGGCAAG	ATCCCCAACT	GGCTTTATCT	TGCAGCCTAC	840
	TTCATAGCCT	ACTTCAACAG	CTGCCTCAAC	GCTGTGATCT	ACGGGCTCCT	CAATGAGAAT	900
10	TTCCGAAGAG	AATACTGGAC	CATCTTCCAT	GCTATGCGGC	ACCCTATCAT	ATTCTTCTCT	960
	GGCCTCATCA	GTGATATTCG	TGAGATGCAG	GAGGCCCGTA	CCCTGGCCCG	CGCCCGTGCC	1020
	CATGCTCGCG	ACCAAGCTCG	TGAACAAGAC	CGTGCCCATG	CCTGTCCTGC	TGTGGAGGAA	1080
	ACCCCGATGA	ATGTCCGGAA	TGTTCCATTA	CCŢGGTGATG	CTGCAGCTGG	CCACCCGAC	1140
* *!	CGTGCCTCTG	GCCACCCTAA	GCCCCATTCC	AGATCCTCCT	CTGCCTATCG	CAAATCTGCC	1200
15	TCTACCCACC	ACAAGTCTGT	CTTTAGCCAC	TCCAAGGCTG	CCTCTGGTCA	CCTCAAGCCT	1260
	GTCTCTGGCC	ACTCCAAGCC	TGCCTCTGGT	CACCCCAAGT	CTGCCACTGT	CTACCCTAAG	1320
	CCTGCCTCTG	TCCATTTCAA	GGCTGACTCT	GTCCATTTCA	AGGGTGACTC	TGTCCATTTC	1380
	AAGCCTGACT	CTGTTCATTT	CAAGCCTGCT	TCCAGCAACC	CCAAGCCCAT	CACTGGCCAC	1440
	CATGTCTCTG	CTGGCAGCCA	CTCCAAGTCT	GCCTTCAGTG	CTGCCACCAG	CCACCCTAAA	1500
20	CCCACCACTG	GCCACATCAA	GCCAGCTACC	AGCCATGCTG	AGCCCACCAC	TGCTGACTAT	1560
	CCCAAGCCTG	CCACTACCAG	CCACCCTAAG	CCCACTGCTG	CTGACAACCC	TGAGCTCTCT	1620
	GCCTCCCATT	GCCCCGAGAT	CCCTGCCATT	GCCCACCCTG	TGTCTGACGA	CAGTGACCTC	1680
	CCTGAGTCGG	CCTCTAGCCC	TGCCGCTGGG	CCCACCAAGC	CTGCTGCCAG	CCAGCTGGAG	1740
	TCTGACACCA	TCGCTGACCT	TCCTGACCCT	ACTGTAGTCA	CTACCAGTAC	CAATGATTAC	1800
. 25	CATGATGTCG	TGGTTGTTGA	TGTTGAAGAT	GATCCTGATG	AAATGGCTGT	GTGA	1854

(219) INFORMATION FOR SEQ ID NO:218:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 617 amino acids
 - (B) TYPE: amino acid

(C)	STRANDEDNESS
-----	--------------

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:218:

	• -		_ #					226		110.2	10.					
5	Met 1	Gly	Pro	Thr	Leu 5	Ala	Val	Pro	Thr	Pro 10	Tyr	Gly	Cys	Ile	Gly 15	Cys
	Lys	Leu	Pro	Gln 20	Pro	Glu	Tyr	Pro	Pro 25	Ala	Leu	Ile	Ile	Phe	Met	Phe
10	Cys	Ala	Met 35	Val	Ile	Thr	Ile	Val 40	Val	Asp	Leu	Ile	Gly 45	Asn	Ser	Met
	Val	Ile 50	Leu	Ala	Val	Thr	Lys 55	Asn	Lys	Lys		60	Asn	Ser	Gly	Asn
,	Ile 65	Phe	Val	Val	Ser	Leu 70	Ser	Val	Ala	Asp	Met 75	Leu	Val	Ala	Ile	Tyr 80
15	Pro	Tyr	Pro	Leu	Met 85	Leu	His	Ala	Met	Ser 90	Ile	Gly	Gly	Trp	Asp 95	Leu
	Ser	Gln	Leu	Gln 100	Cys	Gln	Met	Val	Gly 105	Phe	Ile	Thr	Gly	Leu 110	Ser	Val
20	Val	Gly	Ser 115	Ile	Phe	Asn	Ile	Val 120	Ala	Ile	Ala	Ile	Asn 125	Arg	Tyr	Сув
	Tyr	Ile 130	Cys	His	Ser	Leu	Gln 135	Tyr	Glu	Arg	Ile	Phe 140	Ser	Val	Arg	Asn
	Thr 145	Cys	Ile	Tyr	Leu	Val 150	Ile	Thr	Trp	Ile	Met 155	Thr		Leu		Val
25	Leu	Pro	Asn	Met	Tyr 165	Ile	Gly	Thr	Ile	Glu 170	Tyr	Asp				Tyr
	Thr	Cys	Ile	Phe 180	Asn	Tyr	Leu	Asn	Asn 185	Pro	Val	Phe	Thr	Val 190	Thr	Ile
30	Val	Cys	Ile 195	His	Phe	Val	Leu	Pro 200	Leu	Leu	Ile	Val	Gly 205	Phe	Cys	Tyr
	Val,	Arg 210	Ile	Trp	Thr	Lys	Val 215	Leu	Ala	Ala	Arg	Asp 220	Pro	Ala	Gly	Gln
	Asn 225	Pro	Asp	Asn	Gln	Leu 230	Ala	Glu	Val	Arg	Asn 235	Lys	Leu	Thr	Met	Phe 240
35	Val	Ile	Phe	Leu	Leu 245	Phe	Ala	Val	Cys	Trp 250	Cys	Pro	Ile		Val 255	Leu

	Thr	Val	Leu	Val 260	Ala	Val	Ser	Pro	Lys 265	Glu	Met	Ala	Gly	Lys 270	Ile	Pro
	Asn	Trp	Leu 275	Tyr		Ala	Ala	Tyr 280	Phe	Ile	Ala	Tyr	Phe 285	Asn	Ser	Cys
5	Leu	Asn 290	Ala	Val	Ile	Tyr	Gly 295	Leu	Leu	Asn		Asn 300	Phe	Arg	Arg	Glu
٠.	Tyr 305	Trp		Ile	Phe	His 310	Ala	Met	Arg	His	Pro 315	Ile	Ile	Phe	Phe	Ser 320
10	Gly	Leu	Ile	Ser	Asp 325	Ile	Arg	Glu	Met	Gln 330	Glu	Ala	Arg	Thr	Leu 335	Ala
	Arg	Ala	Arg	Ala 340	His	Ala	Arg	Asp	Gln 345	Ala	Arg	Glu	Gln	Asp 350	Arg	Ala
	His	Ala	Cys 355	Pro	Ala	Val	Glu	Glu 360	Thr	Pro	Met	Asn	Val 365	Arg	Asn	Val
15	Pro	Leu 370	Pro	Gly	Asp	Ala	Ala 375	Ala	Gly	His	Pro	Asp 380	Arg	Ala	Ser	Gly
	His 385	Pro	Lys	Pro	His	Ser 390	Arg	Ser	Ser	Ser	Ala 395	Tyr	Arg	Lys	Ser	Ala 400
20	Ser	Thr	His	His	Lys 405	Ser	Val	Phe	Ser	His 410	Ser	Lys	Ala	Ala	Ser 415	Gly
	His	Leu	Lys	Pro 420	Val	Ser	Gly	His	Ser 425	Lys	Pro	Ala	Ser	Gly 430	His	Pro
	Lys		Ala 435	Thr	Val	Tyr	Pro	Lys 440	Pro	Ala	Ser	Val	His 445	Phe	Lys	Ala
25	Asp	Ser 450		His	Phe	Lys	Gly 455	Asp	Ser	Val	His	Phe 460	Lys	Pro	Asp	Ser
	Val 465	His	Phe	Lys	Pro	Ala 470		Ser	Asn		Lys 475		Ile	Thr		His 480
30	His	Val	Ser	Ala	Gly 485	Ser	His	Ser	Lys	Ser 490		Phe	Ser	Ala	Ala 495	Thr
	Ser	His	Pro	Lys 500	Pro	Thr	Thr	Gly	His 505		Lys	Pro	Ala	Thr 510	Ser	His
	Ala	Glu	Pro 515	Thr	Thr	Ala	Asp	Tyr 520	Pro	Lys	Pro	Ala	Thr 525	Thr	Ser	His
35		530		Thr			535		•			540				
	Pro	Glu	Ile	Pro	Ala	Ile	Ala	His	Pro	Val	Ser	Asp	Asp	Ser	Asp	Leu

	545 550 560
	Pro Glu Ser Ala Ser Ser Pro Ala Ala Gly Pro Thr Lys Pro Ala Ala 565 570 575
5	Ser Gln Leu Glu Ser Asp Thr Ile Ala Asp Leu Pro Asp Pro Thr Val 580 585 590
	Val Thr Thr Ser Thr Asn Asp Tyr His Asp Val Val Val Asp Val 595 600 605
	Glu Asp Asp Pro Asp Glu Met Ala Val 610 615
10	(220) INFORMATION FOR SEQ ID NO:219:
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1548 base pairs (B) TYPE: nucleic acid
15	(C) STRANDEDNESS: single (D) TOPOLOGY: linear
	(ii) MOLECULE TYPE: DNA (genomic)
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:219:
	ATGGGACATA ACGGGAGCTG GATCTCTCCA AATGCCAGCG AGCCGCACAA CGCGTCCGGC 60
	GCCGAGGCTG CGGGTGTGAA CCGCAGCGCG CTCGGGGAGT TCGGCGAGGC GCAGCTGTAC 120
20	CGCCAGTTCA CCACCACCGT GCAGGTCGTC ATCTTCATAG GCTCGCTGCT CGGAAACTTC 180
	ATGGTGTTAT GGTCAACTTG CCGCACAACC GTGTTCAAAT CTGTCACCAA CAGGTTCATT 240
	AAAAACCTGG CCTGCTCGGG GATTTGTGCC AGCCTGGTCT GTGTGCCCTT CGACATCATC 300
	CTCAGCACCA GTCCTCACTG TTGCTGGTGG ATCTACACCA TGCTCTTCTG CAAGGTCGTC 360
	AAATTTTTGC ACAAAGTATT CTGCTCTGTG ACCATCCTCA GCTTCCCTGC TATTGCTTTG 420
25	GACAGGTACT ACTCAGTCCT CTATCCACTG GAGAGGAAAA TATCTGATGC CAAGTCCCGT 480
	GAACTGGTGA TGTACATCTG GGCCCATGCA GTGGTGGCCA GTGTCCCTGT GTTTGCAGTA 540
	ACCAATGTGG CTGACATCTA TGCCACGTCC ACCTGCACGG AAGTCTGGAG CAACTCCTTG 600
	GGCCACCTGG TGTACGTTCT GGTGTATAAC ATCACCACGG TCATTGTGCC TGTGGTGGTG 660
	GTGTTCCTCT TCTTGATACT GATCCGACGG GCCCTGAGTG CCAGCCAGAA GAAGAAGGTC 720
30	ATCATAGCAG CGCTCCGGAC CCCACAGAAC ACCATCTCTA TTCCCTATGC CTCCCAGCGG 780
	GAGGCCGAGC TGAAAGCCAC CCTGCTCTCC ATGGTGATGG TCTTCATCTT GTGTAGCGTG 840
	CCCTATGCCA CCCTGGTCGT CTACCAGACT GTGCTCAATG TCCCTGACAC TTCCGTCTTC 900

	TTGCTGCTCA CTGCTGTTTG GCTGCCCAAA GTCTCCCTGC TGGCAAACCC TGTTCTCTTT	960
	CTTACTGTGA ACAAATCTGT CCGCAAGTGC TTGATAGGGA CCCTGGTGCA ACTACACCAC	1020
	CGGTACAGTC GCCGTAATGT GGTCAGTACA GGGAGTGGCA TGGCTGAGGC CAGCCTGGAA	1080
	CCCAGCATAC GCTCGGGTAG CCAGCTCCTG GAGATGTTCC ACATTGGGCA GCAGCAGATC	1140
5	TTTAAGCCCA CAGAGGATGA GGAAGAGAGT GAGGCCAAGT ACATTGGCTC AGCTGACTTC	1200
	CAGGCCAAGG AGATATTTAG CACCTGCCTG GAGGGAGAGC AGGGGCCACA GTTTGCGCCC	1260
	TCTGCCCCAC CCCTGAGCAC AGTGGACTCT GTATCCCAGG TGGCACCGGC AGCCCCTGTG	1320
	GAACCTGAAA CATTCCCTGA TAAGTATTCC CTGCAGTTTG GCTTTGGGCC TTTTGAGTTG	1380
	CCTCCTCAGT GGCTCTCAGA GACCCGAAAC AGCAAGAAGC GGCTGCTTCC CCCCTTGGGC	1440
10	AACACCCCAG AAGAGCTGAT CCAGACAAAG GTGCCCAAGG TAGGCAGGGT GGAGCGGAAG	1500
	ATGAGCAGAA ACAATAAAGT GAGCATTTTT CCAAAGGTGG ATTCCTAG	1548
	(221) INFORMATION FOR SEQ ID NO:220:	j.
	(i) SEQUENCE CHARACTERISTICS:	
. 15	(A) LENGTH: 515 amino acids (B) TYPE: amino acid	.1
٠.	<pre>(C) STRANDEDNESS: (D) TOPOLOGY: not relevant</pre>	77.
	(ii) MOLECULE TYPE: protein	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:220:	
20	Met Gly His Asn Gly Ser Trp Ile Ser Pro Asn Ala Ser Glu Pro His	5
• .	1 5 10 15	
	Asn Ala Ser Gly Ala Glu Ala Ala Gly Val Asn Arg Ser Ala Leu Gly 20 25 30	7 .
0.5	Glu Phe Gly Glu Ala Gln Leu Tyr Arg Gln Phe Thr Thr Val Glr	1
25	35 40 45	
	Val Val Ile Phe Ile Gly Ser Leu Leu Gly Asn Phe Met Val Leu Trp 50 55 60	Þ
	Ser Thr Cys Arg Thr Thr Val Phe Lys Ser Val Thr Asn Arg Phe Ile	=
	65 70 75 80	
30	Lys Asn Leu Ala Cys Ser Gly Ile Cys Ala Ser Leu Val Cys Val Pro 85 90 95	•
	Phe Asp Ile Ile Leu Ser Thr Ser Pro His Cys Cys Trp Trp Ile Tyr	5
-	100 105 110	

	*:	Thr	met	115	Pne	cys	ьys	vai	120	ьys	·Pne	Leu	HIS	125	vaı	Pne	Cys
		Ser	Val 130	Thr	Ile	Leu	Ser	Phe 135	Pro	Ala	Ile	Ala	Leu 140	Asp	Arg	Tyr	Ту
5		Ser 145	Val	Leu	Tyr	Pro	Leu 150	Glu	Arg	Lys	Ile	Ser 155	Asp	Ala	Lys	Ser	Arg 160
		Glu	Leu	Val	Met	Tyr 165	Ile	Trp	Ala	His	Ala 170	Val	Val	Ala	Ser	Val 175	Pro
10		Val	Phe	Ala	Val 180	Thr	Asn	Val	Ala	Asp 185	Ile	Tyr	Ala	Thr	Ser 190	Thr	Cys
		Thr	Glu	Val 195	Trp	Ser	Asn	Ser	Leu 200	Gly	His	Leu	Val	Tyr 205	Val	Leu	Va]
		Tyr	Asn 210	Ile	Thr	Thr	Val	Ile 215	Val	Pro	Val	Val	Val 220	Val	Phe	Leu	Phe
15		Leu 225	Ile	Leu	Ile	Arg	Arg 230	Ala	Leu	Ser	Ala	Ser 235	Gln	Lys	Lys	Lys	Va]
		Ile	Ile	Ala	Ala	Leu 245	Arg	Thr	Pro	Gln	Asn 250	Thr	Ile	Ser	Ile	Pro 255	Туі
20		Ala	Ser	Gln	Arg 260	Glu	Ala	Glu	Leu	Lys 265	Ala	Thr	Leu	Leu	Ser 270	Met	Va]
		Met	Val	Phe 275	Ile	Leu	Cys	Ser	Val 280	Pro	Tyr	Ala	Thr	Leu 285	Val	Val	Туз
		Gln	Thr 290	Val	Leu	Asn	Val	Pro 295	Asp	Thr	Ser	Val	Phe 300	Leu	Leu	Leu	Thi
25		Ala 305	Val	Trp	Leu	Pro	Lys 310	Val	Ser	Leu	Leu	Ala 315	Asn	Pro	Val	Leu	Phe 320
		Leu	Thr	Val	Asn	Lys 325	Ser	Val	Arg	Lys	Cys 330	Leu	Ile	Gly	Thr	Leu 335	Va]
30	·		Leu		340					345				•	350		
		Gly	Met	Ala 355	Glu	Ala	Ser	Leu	Glu 360	Pro	Ser	Ile	Arg	Ser 365	Gly	Ser	Glı
		Leu	Leu 370	Glu	Met	Phe	His	Ile 375	Gly	Gln	Gln	Gln	Ile 380	Phe	Lys	Pro	Thi
35		385	Asp				390					395					400
		Gln	Ala	Lys	Glu	Ile	Phe	Ser	Thr	Cys	Leu	Glu	Gly	Glu	Gln	Gly	Pro

השפרוכנית חשי תויחתופתו

WO 00/22129 PCT/US99/23938

	• •					405	. • .:.	. •	. •		410					415		
		Gln	Phe	Ala	Pro 420	Ser	Ala	Pro	Pro	Leu 425	Ser	Thr	Val	Asp	Ser 430	Val	Ser	
5	-	Gln	Val	Ala 435	Pro	Ala	Ala	Pro				Glu		Phe 445	Pro	Asp	Lys	
		Tyr	Ser 450	Leu	Gln	Phe	Gly	Phe 455	Gly	Pro	Phe	Glu	Leu 460	Pro	Pro	Gln	Trp	
		Leu 465	Ser	Glu	Thr		Asn 470		Lys	Lys	Arg		Leu			Leu	Gly 480	
10		Asn	Thr	Pro	Glụ	Glu 485	Leu	Ile	Gln	Thr	Lys 490	Val	Pro	Lys	Val	Gly 495	Arg	
-	٠.	Val	Glu	Arg	Lys 500	Met		Arg		Asn 505		Val			Phe 510	Pro	Lys	
15		Val	Asp	Ser 515														٠.
	(222)) INF	ORMA	OITA	1 FOF	SEÇ	Q ID	NO:2	221:									
		(i)	(A)	LEN	CE CH NGTH: PE: r	116	54 ba	ase p		3								
20			(C)	ST	RANDE	EDNES	SS: 8	ingl	le						•			
		(ii) MC	DLECT	LE I	YPE:	DN7	4 (ge	enomi	ic)					· :			
		(xi) SI	EQUE	ICE I	ESCI	RIPTI	ON:	SEQ	ID 1	10:22	21:						
	ATGA	ATCGG	C AC	CATO	CTGC	GG#	TCAC	CTTT	CTG	CAAAE	TAG A	ACAAC	SAAGA	AA CI	GCTG	TGT	}	60
25	TTCCC	GAGAI	G AC	CTTC	ATTGO	CA	AGGT	TTG	CCGC	CGG1	GT 1	rggg	CTG	A GT	CATT	CTTI	: :	120
	GGGC	TTCTG	G GC	CAATO	GCCI	TGC	CCTC	STGG	ATTT	TCTC	TT T	CCAC	CCTCA	AA GI	гссто	GAA		180
	TCCAC	GCCGG	A TI	TTTC	CTGTI	CAZ	ACCTO	GCA	GTAC	GCTG <i>I</i>	ACT I	TCT	CTG	T C	TCTC	CCT	3	240
	CCGT	rcgte	A TO	GACT	TACTA	TGI	rgcgc	GCGT	TCAC	SACTO	GA A	GTTT	regee	A C	TCCC	TTGC	: 3	300
	CGGCT	rggte	C TO	CTTCA	ATGTI	TGC	CATO	BAAC	CGCC	CAGGG	CA C	CATO	CATCI	T C	CTCAC	GGTG	3 3	360
30	GTGG	CGGTA	AG AC	CAGGI	TTTAT	ccc	GGT	GTC	CATO	CCCC	ACC A	ACGCC	CTG	A CA	AGAI	CTCC	:. 4	420
	AATTO	GGACA	AG CA	AGCC	ATCAT	CTC	TTGC	CCTT	CTGT	rggg	CA 1	CACI	GTT	G CC	TAAC	AGTO	2 4	480
	CACC	rccte	a Ac	SAAGA	AGTI	GC1	rgato	CAG	AATO	GCCC	TG C	CAAAC	GTGI	G C	ATCAG	CTTC	: 5	540
	AGCAT	rctgo	C A	racci	TCC	GTO	GCAC	GAA	GCT	TGTI	rcc 1	CCT	GAGI	T CC	CTCCI	GCCC	: 6	600

	CTGGGCAT	CA T	CCTG	TTCT	G CI	CAGC	CAGA	LTA A	ATCI	GGA	GCCI	GCGG	CA C	EAGAC	CAAA	rg	660
	GACCGGCA'	TG C	CAAG	ATCA	A GA	GAGC	CAAA	ACC	TTCA	TCA	TGGI	GGT	GC C	CATCG	TCTI	T	720
٠	GTCATCTG	CT T	CCTT	CCCA	G CG	TGGI	TGTG	CGG	ATCO	:GCA	TCTI	CTGG	CT C	CTGC	ACAC	T	780
	TCGGGCAC	GC A	GAAT	TGTG	A AG	TGTA	'CCG'C	ŢCG	GTGG	ACC	TGGC	GTTC	TŢ 1	TATCA	CTCI	'C	840
5	AGCTTCAC	CT A	CATG	AACA	G CA	TGCT	GGAC	ccc	GTGG	TGT	ACTA	CTTC	TC C	AGCC	CATC	:C	900
•	TTTCCCAA	CT TO	CTTC	TCCA	C TT	TGAT	CAAC	CGC	TGCC	TCC	AGAG	GAAG	AT G	ACAG	GTGA	.G	960
	CCAGATAA	TA AC	CCGC.	AGCA	C GA	.GCGT	CGAG	CTC	ACAG	GGG	ACCC	CAAC	AA A	ACCA	GAGG	C 1	020
	GCTCCAGA	GG CC	STTA	ATGG	C CA	ACTC	CGGT	GAG	CCAT	GGA	GCCC	CTCT	TA T	CTGG	GCCC	'A 1	080
	ACCTCAAA	ra ac	CAT	TCCA	A GA	AGGG	ACAT	TGT	CACC	AAG	AACC	AGCA	TC T	CTGG	AGAA	A 1	140
10	CAGTTGGG	CT GI	rtgc	ATCG/	A GT	AA.			÷				. •	÷		1	164
	(223) INI	FORM	ATIO	N FOI	R SE	Q ID	NO:	222:									
15		(A) (B) (C) (D)	LEI TYI STI TOI	CE CH NGTH: PE: & RANDE POLOC ULE 1	: 38 amin EDNE EY: :	7 am o ac SS: not	ino id rele	acid vant	s								
	(xi	.) SE	QUE	NCE I	ESC	RIPT	ION:	SEQ	ID :	NO:2	22:						
20	Met 1	Asn	Arg	His	His 5	Leu	Gln	Asp	His	Phe	Leu	Glu	Ιĺε	Asp	Lys 15	Lys	
	Asn	Cys	Cys	Val 20	Phe	Arg	Asp	Asp	Phe 25		Ala	_			Pro	Pro	
	Val		Gly 35	Leu	Glu	Phe	Ile	Phe 40	Gly	Leu	Leu	Gly	Asn 45	Gly	Leu	Ala	
25	Leu	Trp 50	Ile	Phe	Cys	Phe	His 55	Leu	Lys	Ser	Trp	Lys 60	Ser	Ser	Arg	Ile	
	Phe 65	Leu	Phe	Asn	Leu	Ala 70	Val	Ala	Asp	Phe	Leu 75	Leu	Ile	Ile	Cys	Leu 80	
30	Pro	Phe	Val	Met	Asp 85	Tyr	Tyr	Val	Arg	Arg 90	Ser	Asp	Trp	Lys	Phe 95	Gly	
	Asp	Ile	Pro	Cys 100	Arg	Leu	Val	Leu	Phe 105	Met	Phe	Ala	Met	Asn 110	Arg	Gln	
	Gly		Ile	Ile	Phe	Leu	Thr	Val	Val	Ala	Val	Asp	Arg	Tyr	Phe	Arg	

		Val	Val 130	His	Pro	Hiś	His	Ala 135	Leu	Asn	Ĺys	Ile	Ser 140	Asn	Trp	Thr	Ala
		Ala 145	Ile	Ile	Ser	Cys	Leu 150	Leu	Trp	Gly	Ile	Thr 155	Val	Gly	Leu	Thr	Val 160
5		His	Leu	Leu	Lys	Lys 165	Lys	Leu	Leu	Ile	Gln 170	Asn	Gly	Pro	Ala	Asn 175	Val
	• **	Cys	Ile	Ser	Phe 180	Ser	Ile	Cys	His	Thr 185	Phe	Arg	Trp	His	Glu 190	Ala	Met
10	***	Phe	Leu	Leu 195	Glu	Phe	Leu	Leu	Pro 200	Leu	Gly	Ile	Ile	Leu 205	Phe	Cys	Ser
		Ala	Arg 210	Ile	Ile	Trp	Ser	Leu 215	Arg	Gln	Arg	Gln	Met 220	Asp	Arg	His	Ala
:		Lys 225	Ile	Lys	Arg	Ala	Lys 230	Thr	Phe	Ile	Met	Val 235	Val	Ala	Ile :	Val	Phe 240
15		Val	Ile	Cys	Phe	Leu 245	Pro	Ser	Val	Val	Val 250	Arg	Ile	Arg	Ile	Phe 255	Trp
		Leu	Leu	His	Thr 260	Ser	Gly	Thr	Gln	Asn 265	Cys	Glu	Val	Tyr	Arg 270	Ser	Val
20		Asp	Leu	Ala 275	Phe	Phe	Ile	Thr	Leu 280	Ser	Phe	Thr	Tyr	Met 285	Asn	Ser	Met
		Leu	Asp 290	Pro	Val	Val	Tyr	Tyr 295	Phe	Ser	Ser	Pro	Ser 300	Phe	Pro	Asn	Phe
		Phe 305	Ser	Thr	Leu	Ile	Asn 310	Arg	Cys	Leu	Gln	Arg 315	Lys	Met	Thr	Gly	Glu 320
25		Pro	Asp	Asn	Asn	Arg 325	Ser	Thr	Ser	Val	Glu 330	Leu	Thr	Gly	Asp	Pro 335	Asn
		Lys	Thr	Arg	Gly 340	Ala	Pro	Glu	Ala	Leu 345	Met	Ala	Asn	Ser	Gly 350	Glu	Pro
30		Trp	Ser	Pro 355	Ser	Tyr	Leu	Gly	Pro 360	Thr	Ser	Asn	Asn	His 365	Ser	Lys	Lys
		Gly	His 370	Cys	His	Gln	Glu	Pro 375	Ala	Ser	Leu	Glu	Lys 380	Gln	Leu	Gly	Cys
		Cys 385	Ile	Glu													
35	(224)) INI	FORM	ATIO	v FOI	R SE	Q ID	NO:	223:								

- 35 (224) INFORMATION FOR SEQ ID NO:223:
 - (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1212 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: DNA (genomic)
- 5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:223:

		ATGGCTTGCA	ATGGCAGTGC	GGCCAGGGGG	CACTTTGACC	CTGAGGACTT	GAACCTGACT	60
		GACGAGGCAC	TGAGACTCAA	GTACCTGGGG	CCCCAGCAGA	CAGAGCTGTT	CATGCCCATC	120
		TGTGCCACAT	ACCTGCTGAT	CTTCGTGGTG	GGCGCTGTGG	GCAATGGGCT	GACCTGTCTG	180
		GTCATCCTGC	GCCACAAGGC	CATGCGCACG	CCTACCAACT	ACTACCTCTT	CAGCCTGGCC	240
	10	GTGTCGGACC	TGCTGGTGCT	GCTGGTGGGC	CTGCCCCTGG	AGCTCTATGA	GATGTGGCAC	300
		AACTACCCCT	TCCTGCTGGG	CGTTGGTGGC	TGCTATTTCC	GCACGCTACT	GTTTGAGATG	360
		GTCTGCCTGG	CCTCAGTGCT	CAACGTCACT	GCCCTGAGCG	TGGAACGCTA	TGTGGCCGTG	420
		GTGCACCCAC	TCCAGGCCAG	GTCCATGGTG	ACGCGGGCCC	ATGTGCGCCG	AGTGCTTGGG	480
		GCCGTCTGGG	GTCTTGCCAT	GCTCTGCTCC	CTGCCCAACA	CCAGCCTGCA	CGGCATCCGG	540
,	15	CAGCTGCACG	TGCCCTGCCG	GGGCCCAGTG	CCAGACTCAG	CTGTTTGCAT	GCTGGTCCGC	600
P.		CCACGGGCCC	TCTACAACAT	GGTAGTGCAG	ACCACCGCGC	TGCTCTTCTT	СТССССССС	660
		ATGGCCATCA	TGAGCGTGCT	CTACCTGCTC	ATTGGGCTGC	GACTGCGGCG	GGAGAGGCTG	720
		CTGCTCATGC	AGGAGGCCAA	GGGCAGGGGC	TCTGCAGCAG	CCAGGTCCAG	ATACACCTGC	780
		AGGCTCCAGC	AGCACGATCG	GGGCCGGAGA	CAAGTGAAGA	AGATGCTGTT	TGTCCTGGTC	840
	20	GTGGTGTTTG	GCATCTGCTG	GGCCCCGTTC	CACGCCGACC	GCGTCATGTG	GAGCGTCGTG	900
		TCACAGTGGA	CAGATGGCCT	GCACCTGGCC	TTCCAGCACG	TGCACGTCAT	CTCCGGCATC	960
		TTCTTCTACC	TGGGCTCGGC	GGCCAACCCC	GTGCTCTATA	GCCTCATGTC	CAGCCGCTTC	1020
		CGAGAGACCT	TCCAGGAGGC	CCTGTGCCTC	GGGGCCTGCT	GCCATCGCCT	CAGACCCCGC	1080
		CACAGCTCCC	ACAGCCTCAG	CAGGATGACC	ACAGGCAGCA	CCCTGTGTGA	TGTGGGCTCC	1140
	25	CTGGGCAGCT	GGGTCCACCC	CCTGGCTGGG	AACGATGGCC	CAGAGGCGCA	GCAAGAGACC	1200
		GATCCATCCT	GA					1212

(225) INFORMATION FOR SEQ ID NO:224:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 403 amino acids
- 0 (B) TYPE: amino acid

ANSULUT - MU MUSTISMI I -

(C)	STRANDEDNES	s:
-----	-------------	----

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:224:

	•	•	_					-								
5	Met 1	Ala	Cys	Asn	Gly 5	Ser	Ala	Ala	Arg	Gly 10	His	Phe	Asp	Pro	Glu 15	Asp
	Leu	Asn	Leu	Thr 20	Asp	Glu	Ala	Leu	Arg 25	Leu	Lys	Tyr	Leu	Gly 30	Pro	Gln
10	Gln	Thr	Glu 35	Leu	Phe	Met	Pro	Ile 40	Cys	Ala	Thr	Tyr	Leu 45	Leu	Ile	Phe
	Val	Val 50	Gly	Ala	Val	Gly	Asn 55	Gly	Leu	Thr	Cys	Leu 60	Val	Ile	Leu	Arg
• . :	His 65	Lys	Ala	Met	Arg	Thr 70	Pro	Thr	Asn	Tyr	Tyr 75	Leu	Phe	Ser	Leu	Ala 80
15	Val	Ser	Asp	Leu	Leu 85	Val	Leu	Leu	Val	Gly 90	Leu	Pro	Leu	Glu	Leu 95	Tyr
	Glu	Met	Trp	His 100	Asn	Tyr	Pro	Phe	Leu 105	Leu	Gly	Val	Gly	Gly 110	Cys	Tyr
20	Phe	Arg	Thr 115	Leu	Leu	Phe	Glu	Met 120	Va]	Cys	Leu	Ala	Ser 125	Val	Leu	Asn
	Val	Thr 130	Ala	Leu	Ser	Val	Glu 135	Arg	Tyr	Val	Ala	Val 140	Val	His	Pro	Leu
	Gln 145	Ala	Arg	Ser	Met	Val 150	Thr	Arg	Ala	His	Val 155	Arg	Arg	Val	Leu	Gly 160
25	Ala	Val	Trp	Gly	Leu 165	Ala	Met	Leu	Cys	Ser 170	Leu	Pro	Asn	Thr	Ser 175	Leu
	His	Gly	Ile	Arg 180	Gln	Leu	His	Val	Pro 185	Cys	Arg	Gly	Pro	Val 190	Pro	Asp
30	Ser	Ala	Val 195	Cys	Met	Leu	Val	Arg 200	Pro	Arg	Ala	Leu	Tyr 205	Asn	Met	Val
	Val	Gln 210	Thr	Thr	Ala	Leu	Leu 215	Phe	Phe	Cys	Leu	Pro 220	Met	Ala	Ile	Met
	Ser 225	Val	Leu	Tyr	Leu	Leu 230	Ile	Gly	Leu	Arg	Leu 235	Arg	Arg	Glu	Arg	Leu 240
35	Leu	Leu	Met	Gln	Glu 245	Ala	Lys	Gly	Arg	Gly 250	Ser	Ala	Ala	Ala	Arg 255	Ser

	Arg	Tyr.		Cys 260	Arg	Leu	Gln	Gln	His 265	Asp	Arg	Gly	Arg	Arg 270	Gln	Val	
,	Lys	Lys i	Met : 275	Leu	Phe	Val	Leu	Val 280		Val	Phe	Gly	Ile 285	Cys	Trp	Ala	
5	Pro	Phe 1 290	His .	Ala	Asp	Arg	Val 295	Met	Trp	Ser	Val	Val 300	Ser	Gln	Trp	Thr	
	Asp 305	Gly:	Leu 1	His	Leu	Ala 310	Phe	Gln	His	Val	His 315	Val	Ile	Ser	Gly	Ile 320	
0	Phe	Phe '	Tyr :		Gly 325	Ser	Ala	Ala	Asn	Pro 330	Val	Leu	Tyr	Ser	Leu 335	Met	
	Ser	Ser /		Phe 340	Arg	Glu	Thr	Phe	Gln 345	Glu	Ala	Leu	Cys		Gly	Ala	
	Cys	Cys :	His : 355	Arg	Leu	Arg	Pro	Arg 360	His	Ser	Ser	His	Ser 365	Leu	Ser	Arg	
5	Met	Thr '	Thr	Gly	Ser	Thr	Leu 375	Cys	Asp	Val	Gly	Ser 380	Leu	Gly	Ser	Trp	
	Val 385	His	Pro :	Leu	Ala	Gly 390	Asn	Asp	Gly	Pro	Glu 395	Ala	Gln	Gln	Glu	Thr 400	
20	Asp	Pro	Ser						•				•				
	(226) INF	FORMA	TION	FOR	SEÇ) ID	NO:2	25:									
25	(i)	(B) (C)	LEN TYP STR	GTH: E: n ANDE	109 ucle DNES	98 ba eic a SS: s	se p cid sing]	airs						. •			
	/::	i) MO	TOP					nomi	.: !a\		•						
		i) SE								: : :) 5 .						
	ATGGGGAAG											CGACO	a TA	ACCAT	rccac	2	6
30	CAGACGCTO																12
	CTGTCCCTC	CT AC	TTCG	GCTA	CCI	rgcao	FATC	AAG	3CCC	GA 1	CGA	CTG	G C	GTGT?	ACCTO	3	18
	TGCAACCT	GA CG	GTGG	CCGA	. cci	CTT	CTAC	ATC	rgcto	GC :	rgcco	CTTCT	rg go	CTGC	AGTAC	2	24
	GTGCTGCAG	GC AC	GACA	ACTG	GTO	CTCA	CGGC	GAC	CTGT	CCT (GCCA	GTG	rg co	GCA	CCT	2	30
	CTGTACGAC	GA AC	ATCT	'ACAT	CAC	GCGT	GGC	TTC	CTCTC	GCT (CAT	CTCC	et Go	SACC	CTAC	2	36
35	CTGGCTGTC	G CC	CATC	CCTT	cco	CTT	CCAC	CAG	rtcc	GA (CCT	BAAG	C G	GCCG:	rcgg	2	42

WO 00/22129 PCT/US99/23938

				191			
•	GTCAGCGTGG	TCATCTGGGC	CAAGGAGCTG	CTGACCAGCA	TCTACTTCCT	GATGCACGAG	480
	GAGGTCATCG	AGGACGAGAA	CCAGCACCGC	GTGTGCTTTG	AGCACTACCC	CATCCAGGCA	540
	TGGCAGCGCG	CCATCAACTA	CTACCGCTTC	CTGGTGGGCT	TCCTCTTCCC	CATCTGCCTG	600
	CTGCTGGCGT	CCTACCAGGG	CATCCTGCGC	GCCGTGCGCC	GGAGCCACGG	CACCCAGAAG	660
5	AGCCGCAAGG	ACCAGATCAA	GCGGCTGGTG	CTCAGCACCG	TGGTCATCTT	CCTGGCCTGC	720
	TTCCTGCCCT	ACCACGTGTT	GCTGCTGGTG	CGCAGCGTCT	GGGAGGCCAG	CTGCGACTTC	780
	GCCAAGGGCG	TTTTCAACGC	CTACCACTTC	TCCCTCCTGC	TCACCAGCTT	CAACTGCGTC	840
	GCCGACCCCG	TGCTCTACTG	CTTCGTCAGC	GAGACCACCC	ACCGGGACCT	GGCCCGCCTC	900
	CGCGGGGCCT	GCCTGGCCTT	CCTCACCTGC	TCCAGGACCG	GCCGGGCCAG	GGAGGCCTAC	960
10	CCGCTGGGTG	CCCCCGAGGC	CTCCGGGAAA	AGCGGGGCCC	AGGGTGAGGA	GCCCGAGCTG	1020
	TTGACCAAGC	TCCACCCGGC	CTTCCAGACC	CCTAACTCGC	CAGGGTCGGG	CGGGTTCCCC	1080
	ACGGGCAGGT	TGGCCTAG					1098
	(227) · INFO	RMATION FOR	SEQ ID NO:2	226 :			
15	(SEQUENCE CHA (A) LENGTH: (B) TYPE: an (C) STRANDEL	365 amino a nino acid				
		(D) TOPOLOGY	/· not relea	rant			

- (ii) MOLECULE TYPE: protein
- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:226:

Met Gly Asn Ile Thr Ala Asp Asn Ser Ser Met Ser Cys Thr Ile Asp

His Thr Ile His Gln Thr Leu Ala Pro Val Val Tyr Val Thr Val Leu 20

25 Val Val Gly Phe Pro Ala Asn Cys Leu Ser Leu Tyr Phe Gly Tyr Leu

Gln Ile Lys Ala Arg Asn Glu Leu Gly Val Tyr Leu Cys Asn Leu Thr

Val Ala Asp Leu Phe Tyr Ile Cys Ser Leu Pro Phe Trp Leu Gln Tyr 30 65

Val Leu Gln His Asp Asn Trp Ser His Gly Asp Leu Ser Cys Gln Val

Cys Gly Ile Leu Leu Tyr Glu Asn Ile Tyr Ile Ser Val Gly Phe Leu

100 105 110

		Cys	Cys	Ile 115	Ser	Val	Asp	Arg	Tyr 120	Leu	Ala	Val	Ala	His 125		Phe	Arg
5		Phe	His 130	Gln	Phe	Arg	Thr	Leu 135	Lys	Ala	Ala	Val	Gly 140	Val	Ser	Val	Val
		Ile 145	Trp	Ala	Lys	Glu	Leu 150	Leu	Thr	Ser	Ile	Tyr 155	Phe	Leu	Met	His	Glu 160
		Glu	Val	Ile	Glu	Asp 165	Glu	Asn	Gln	His	Arg 170	Val	Cys	Phe	Glu	His 175	Tyr
10		Pro	Ile	Gln	Ala 180	Trp	Gln	Arg	Ala	Ile 185	Asn	Tyr	Tyr	Arg	Phe 190	Leu	Val
		Gly	Phe	Leu 195	Phe	Pro	Ile	Cys	Leu 200	Leu	Leu	Ala	Ser	Tyr 205	Gln	Gly	Ile
15		Leu	Arg 210	Ala	Val	Arg	Arg	Ser 215	His	Gly	Thr	Gln	Lys 220	Ser	Arg	Lys	Asp
		Gln 225	Ile	Lys	Arg	Leu	Val 230	Leu	Ser	Thr	Val	Val 235	Ile	Phe	Leu	Ala	Cys 240
		Phe	Leu	Pro	Tyr	His 245	Val	Leu	Leu		Val 250	Arg	Ser	Val	Trp	Glu 255	
20		Ser	Cys	Asp	Phe 260	Ala	Lys	Gly	Val	Phe 265	Asn	Ala	Tyr	His	Phe 270	Ser	Leu
		Leu	Leu	Thr 275	Ser	Phe	Asn	Cys	Val 280	Ala	Asp	Pro	Val	Leu 285	Tyr	Cys	Phe
25		Val	Ser 290	Glu	Thr	Thr	His	Arg 295	Asp	Leu	Ala	Arg	Leu 300		Gly	Ala	Cys
		Leu 305	Ala	Phe	Leu	Thr	Cys 310		Arg	Thr	Gly	Arg 315		Arg	Glu	Ala	Tyr 320
		Pro	Leu	Gly	Ala	Pro 325	Glu	Ala	Ser	Gly	Lys 330	Ser	Gly	Ala	Gln	Gly 335	Glu
30		Glu	Pro	Glu	Leu 340		Thr	Lys		His 345	Pro	Ala	Phe	Gln	Thr 350	Pro	Asn
		Ser	Pro	Gly 355		Gly	Gly	Phe	Pro 360		Gly	Arg	Leu	Ala 365			
	(228) IN	FORM	ATIO	n fo	R SE	Q ID	NO:	227:								
35		(i		QUEN						S	•	•					

(B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:227:

5	ATGGATATTC	TTTGTGAAGA	AAATACTTCT	TTGAGCTCAA	CTACGAACTC	CCTAATGCAA	60
	TTAAATGATG	ACAACAGGCT	CTACAGTAAT	GACTTTAACT	CCGGAGAAGC	TAACACTTCT	120
	GATGCATTTA	ACTGGACAGT	CGACTCTGAA	AATCGAACCA	ACCTTTCCTG	TGAAGGGTGC	180
	CTCTCACCGT	CGTGTCTCTC	CTTACTTCAT	CTCCAGGAAA	AAAACTGGTC	TGCTTTACTG	240
	ACAGCCGTAG	TGATTATTCT	AACTATTGCT	GGAAACATAC	TCGTCATCAT	GGCAGTGTCC	300
10	CTAGAGAAAA	AGCTGCAGAA	TGCCACCAAC	TATTTCCTGA	TGTCACTTGC	CATAGCTGAT	360
	ATGCTGCTGG	GTTTCCTTGT	CATGCCCGTG	TCCATGTTAA	CCATCCTGTA	TGGGTACCGG	420
	TGGCCTCTGC	CGAGCAAGCT	TTGTGCAGTC	TGGATTTACC	TGGACGTGCT	CTTCTCCACG	480
	GCCTCCATCA	TGCACCTCTG	CGCCATCTCG	CTGGACCGCT	ACGTCGCCAT	CCAGAATCCC	540
	ATCCACCACA	GCCGCTTCAA	CTCCAGAACT	AAGGCATTTC	TGAAAATCAT	TGCTGTTTGG	600
15	ACCATATCAG	TAGGTATATC	CATGCCAATA	CCAGTCTTTG	GGCTACAGGA	CGATTCGAAG	660
	GTCTTTAAGG	AGGGGAGTTG	CTTACTCGCC	GATGATAACT	TTGTCCTGAT	CGGCTCTTTT	720
	GTGTCATTTT	TCATTCCCTT	AACCATCATG	GTGATCACCT	ACTTTCTAAC	TATCAAGTCA	780
	CTCCAGAAAG	AAGCTACTTT	GTGTGTAAGT	GATCTTGGCA	CACGGGCCAA	ATTAGCTTCT	840
	TTCAGCTTCC	TCCCTCAGAG	TTCTTTGTCT	TCAGAAAAGC	TCTTCCAGCG	GTCGATCCAT	900
20	AGGGAGCCAG	GGTCCTACAC	AGGCAGGAGG	ACTATGCAGT	CCATCAGCAA	TGAGCAAAAG	960
	GCAAAGAAGG	TGCTGGGCAT	CGTCTTCTTC	CTGTTTGTGG	TGATGTGGTG	CCCTTTCTTC	1020
	ATCACAAACA	TCATGGCCGT	CATCTGCAAA	GAGTCCTGCA	ATGAGGATGT	CATTGGGGCC	1080
	CTGCTCAATG	TGTTTGTTTG	GATCGGTTAT	CTCTCTTCAG	CAGTCAACCC	ACTAGTCTAC	1140
	ACACTGTTCA	ACAAGACCTA	TAGGTCAGCC	TTTTCACGGT	ATATTCAGTG	TCAGTACAAG	1200
25	GAAAACAAAA	AACCATTGCA	GTTAATTTTA	GTGAACACAA	TACCGGCTTT	GGCCTACAAG	1260
	TCTAGCCAAC	TTCAAATGGG	ACAAAAAAG	AATTCAAAGC	AAGATGCCAA	GACAACAGAT	1320
	AATGACTGCT	CAATGGTTGC	TCTAGGAAAG	CAGTATTCTG	AAGAGGCTTC	TAAAGACAAT	13,80
	AGCGACGGAG	TGAATGAAAA	GGTGAGCTGT	GTGTGA			1416

(229) INFORMATION FOR SEQ ID NO:228: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 470 amino acids (B) TYPE: amino acid 5 (C) STRANDEDNESS: (D) TOPOLOGY: not relevant (ii) MOLECULE TYPE: protein (xi) SEQUENCE DESCRIPTION: SEQ ID NO:228: Met Asp Ile Leu Cys Glu Glu Asn Thr Ser Leu Ser Ser Thr Thr Asn 10 ı 10 Ser Leu Met Gln Leu Asn Asp Asp Asn Arg Leu Tyr Ser Asn Asp Phe 25 Asn Ser Gly Glu Ala Asn Thr Ser Asp Ala Phe Asn Trp Thr Val Asp 40 15 Ser Glu Asn Arg Thr Asn Leu Ser Cys Glu Gly Cys Leu Ser Pro Ser 55 Cys Leu Ser Leu Leu His Leu Gln Glu Lys Asn Trp Ser Ala Leu Leu 70 Thr Ala Val Val Ile Ile Leu Thr Ile Ala Gly Asn Ile Leu Val Ile 20 90 Met Ala Val Ser Leu Glu Lys Lys Leu Gln Asn Ala Thr Asn Tyr Phe Leu Met Ser Leu Ala Ile Ala Asp Met Leu Leu Gly Phe Leu Val Met 120 25 Pro Val Ser Met Leu Thr Ile Leu Tyr Gly Tyr Arg Trp Pro Leu Pro 135 Ser Lys Leu Cys Ala Val Trp Ile Tyr Leu Asp Val Leu Phe Ser Thr 145 150 155 Ala Ser Ile Met His Leu Cys Ala Ile Ser Leu Asp Arg Tyr Val Ala 30 165 Ile Gln Asn Pro Ile His His Ser Arg Phe Asn Ser Arg Thr Lys Ala 180 Phe Leu Lys Ile Ile Ala Val Trp Thr Ile Ser Val Gly Ile Ser Met 200 35 Pro Ile Pro Val Phe Gly Leu Gln Asp Asp Ser Lys Val Phe Lys Glu

Gly Ser Cys Leu Leu Ala Asp Asp Asn Phe Val Leu Ile Gly Ser Phe

230

								٠.									
		Val	Ser	Phe	Phe	Ile 245	Pro	Leu	Thr	Ile	Met 250	Val	Ile	Thr	Tyr	Phe 255	Let
		Thr	Ile	Lys	Ser 260	Leu	Gln	Lys	Glu	Ala 265	Thr	Leu	Cys	Val	Ser 270	Asp	Leu
5		Gly	Thr	Arg 275	Ala	Lys	Leu	Ala	Ser 280	Phe	Ser	Phe	Leu	Pro 285	Gln	Ser	Ser
		Leu	Ser 290	Ser	Glu	Lys	Leu	Phe 295	Gln	Arg	Ser	Ile	His 300	Arg	Glu	Pro	Gly
10	•	Ser 305	Tyr	Thr	Gly	Arg	Arg 310	Thr	Met	Gln	Ser	Ile 315	Ser	Asn	Glu	Gln	Lys 320
		Ala	Lys	Lys	Val	Leu 325	Gly	Ile	Val	Phe	Phe 330	Leu	Phe	Val	Val	Met 335	Trp
	٠.,.	Cys	Pro	Phe	Phe 340	Ile	Thr	Asn		Met 345	Ala	Val	Ile	Cys	Lys 350	Glu	Ser
15		Cys	Asn	Glu 355	Asp	Val	Ile	Gly	Ala 360	Leu	Leu	Asn	Val	Phe 365	Val	Trp	Ile
•		Gly ·	Tyr 370	Leu	Ser	Ser	Ala	Val 375	Asn	Pro	Leu	Val	Tyr 380	Thr	Leu	Phe	Asn
20		Lys 385		Tyr	Arg	Ser	Ala 390		Ser	Arg	Tyr	Ile 395	Gln	Cys	Gln	Tyr	Lys 400
		Glu	Asn	Lys	Lys	Pro 405	Leu	Gln	Leu	Ile	Leu 410	Val	Asn	Thr	Ile	Pro 415	Ala
		Leu	Ala	Tyr	Lys 420	Ser	Ser	Gln	Leu	Gln 425	Met	Gly	Gln	Lys	Lys 430	Asn	Ser
25		Lys	Gln	Asp 435	Ala	Lys	Thr	Thr	Asp 440	Asn	Asp	Cys	Ser	Met 445	Val	Ala	Leu
		Gly	Lys 450	Gln	Tyr	Ser		Glu 455	Ala	Ser	Lys	Asp	Asn 460	Ser	Asp	Gly	Val
30	,	Asn 465	Glu	Lys	Val	Ser	Cys 470	Val			٠						
	(230)) INE	FORM	TION	1 FOF	SEÇ	O ID	NO:2	229:								
		(i)	SEC	QUENC	CE CH	IARAC	TER	STIC	CS:								
35			(A) (B) (C)	LEN TYI STI	NGTH: PE: r RANDE	137 nucle EDNES	77 ba eic a SS: s	ase pacid	oairs	5							
	<u>.</u> *		,					-									

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:229:

	ATGGTGAACC	TGAGGAATGC	GGTGCATTCA	TTCCTTGTGC	ACCTAATTGG	CCTATTGGTT	60
	TGGCAATGTG	ATATTTCTGT	GAGCCCAGTA	GCAGCTATAG	TAACTGACAT	TTTCAATACC	120
	TCCGATGGTG	GACGCTTCAA	ATTCCCAGAC	GGGGTACAAA	ACTGGCCAGC	ACTTTCAATC	180
:	5 GTCATCATAA	TAATCATGAC	AATAGGTGGC	AACATCCTTG	TGATCATGGC	AGTAAGCATG	240
	GAAAAGAAAC	TGCACAATGC	CACCAATTAC	TTCTTAATGT	CCCTAGCCAT	TGCTGATATG	300
	CTAGTGGGAC	TACTTGTCAT	GCCCCTGTCT	CTCCTGGCAA	TCCTTTATGA	TTATGTCTGG	360
	CCACTACCTA	GATATTTGTG	CCCCGTCTGG	ATTTCTTTAG	ATGTTTTATT	TTCAACAGCG	420
	TCCATCATGC	ACCTCTGCGC	TATATCGCTG	GATCGGTATG	TAGCAATACG	TAATCCTATT	480
10) GAGCATAGCC	GTTTCAATTC	GCGGACTAAG	GCCATCATGA	AGATTGCTAT	TGTTTGGGCA	540
	ATTTCTATAG	GTGTATCAGT	TCCTATCCCT	GTGATTGGAC	TGAGGGACGA	AGAAAAGGTG	600
	TTCGTGAACA	ACACGACGTG	CGTGCTCAAC	GACCCAAATT	TCGTTCTTAT	TGGGTCCTTC	660
	GTAGCTTTCT	TCATACCGCT	GACGATTATG	GTGATTACGT	ATTGCCTGAC	CATCTACGTT	720
	CTGCGCCGAC	AAGCTTTGAT	GTTACTGCAC	GGCCACACCG	AGGAACCGCC	TGGACTAAGT	780
15	CTGGATTTCC	TGAAGTGCTG	CAAGAGGAAT	ACGGCCGAGG	AAGAGAACTC	TGCAAACCCT	840
	AACCAAGACC	AGAACGCACG	CCGAAGAAAG	AAGAAGGAGA	GACGTCCTAG	GGGCACCATG	900
	CAGGCTATCA	ACAATGAAAG	AAAAGCTAAG	AAAGTCCTTG	GGATTGTTTT	CTTTGTGTTT	960
	CTGATCATGT	GGTGCCCATT	TTTCATTACC	AATATTCTGT	CTGTTCTTTG	TGAGAAGTCC	1020
	TGTAACCAAA	AGCTCATGGA	AAAGCTTCTG	AATGTGTTTG	TTTGGATTGG	CTATGTTTGT	1080
20) TCAGGAATCA	ATCCTCTGGT	GTATACTCTG	TTCAACAAAA	TTTACCGAAG	GGCATTCTCC	1140
,	AACTATTTGC	GTTGCAATTA	TAAGGTAGAG	AAAAAGCCTC	CTGTCAGGCA	GATTCCAAGA	1200
	GTTGCCGCCA	CTGCTTTGTC	TGGGAGGGAG	CTTAATGTTA	ACATTTATCG	GCATACCAAT	1260
	GAACCGGTGA	TCGAGAAAGC	CAGTGACAAT	GAGCCCGGTA	TAGAGATGCA	AGTTGAGAAT	1320
	TTAGAGTTAC	CAGTAAATCC	CTCCAGTGTG	GTTAGCGAAA	GGATTAGCAG	TGTGTGA	1377
	•						

- 25 (231) INFORMATION FOR SEQ ID NO:230:
 - (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 458 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:

(D)	TOPOLOGY:	not	relevant
-----	-----------	-----	----------

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:230:

				_				_	_								
5		Met 1	Val	Asn	Leu	Arg 5	Asn	Ala	Val	His	Ser 10	Phe	Leu	Val	His	Leu 15	Ile
		Gly	Leu	Leu	Val 20	Trp	Gln	Cys	Asp	Ile 25	Ser	Val	Ser	Pro	Val 30	Ala	Ala
		Ile	Val	Thr 35	Asp	Ile	Phe	Asn	Thr 40	Ser	Asp	Gly	Gly	Arg 45	Phe	Lys	Phe
10		Pro	Asp 50	Gly	Val	Gln	Asn	Trp 55	Pro	Ala	Leu	Ser	Ile 60	Val	Ile	Ile	Ile
		lle 65	Met	Thr	Ile	Gly	Gly 70	Asn	Ile	Leu	Val	Ile 75	Met	Ala	Val	Ser	Met 80
15	•	Glu	Lys	Lys	Leu	His 85	Asn	Ala	Thr	Asn	Tyr 90	Phe	Leu	Met	Ser	Leu 95	Ala '
		Ile	Ala	Asp	Met 100	Leu	Val	Gly	Leu	Leu 105	Val	Met	Pro	Leu	Ser 110	Leu	Leu
		Ala	Ile	Leu 115	Tyr	Asp	Tyr	Val	Trp 120	Pro	Leu	Pro	Arg	Tyr 125	Leu	Cys	Pro
20		Val	Trp 130	Ile	Ser	Leu	Asp	Val 135	Leu	Phe	Ser	Thr	Ala 140	Ser	Ile	Met	His
		Leu 145	Cys	Ala	Ile	Ser	Leu 150	Asp	Arg	Tyr	Val	Ala 155	Ile	Arg	Asn	Pro	Ile 160
25	•	Glu	His	Ser	Arg	Phe 165	Asn	Ser	Arg	Thr	Lys 170	Ala	Ile	Met	Lys	Ile 175	Ala
		Ile	Val	Trp	Ala 180	Ile	Ser	Ile	Gly	Val 185	Ser	Val	Pro	Ile	Pro 190	Val	Ile
		Gly	Leu	Arg 195	Asp	Glu	Glu	Lys	Val 200	Phe	Val	Asn	Asn	Thr 205	Thr	Cys	Val
30		Leu	Asn 210	Asp	Pro	Asn	Phe	Val 215	Leu	Ile	Gly	Ser	Phe 220	Val	Ala	Phe	Phe
	,	Ile 225	Pro	Leu	Thr	Ile	Met 230	Val	Ile	Thr	Tyr	Cys 235	Leu	Thr	Ile	Tyr	Val 240
35		Leu	Arg	Arg	Gln	Ala 245	Leu	Met	Leu	Leu	His 250	Gly	His	Thr	Glu	Glu 255	Pro
		Pro	Gly	Leu	Ser	Leu	Asp	Phe	Leu	Lys	Cys	Cys	Lys	Arg	Asn	Thr	Ala

				260				٠	265		٠			270		
	Glu	Glu	Glu 275	Asn	Ser	Ala	Asn	Pro 280	Asn	Gln	Asp	Gln	Asn 285	Ala	Arg	Arg
5	Arg	Lys 290	Lys	Lys	Glu	Arg	Arg 295	Pro	Arg	Gly	Thr	Met 300	Gln	Ala	Ile	Asn
	Asn 305	Glu	Arg	Lys	Ala	Lys 310	Lys	Val	Leu	Gly	Ile 315	Val	Phe	Phe	Val	Phe 320
	Leu	Ile	Met	Trp	Cys 325	Pro	Phe	Phe	Ile	Thr 330	Asn	Ile	Leu	Ser	Val 335	Leu
10	Cys	Glu	Lys	Ser 340	Cys	Asn	Gln	Lys	Leu 345	Met	Glu	Lys	Leu	Leu 350	Asn	Val
	Phe	Val	Trp 355	Ile	Gly	Tyr	Val	Cys 360	Ser	Gly	Ile	Asn	Pro 365.		Val	Tyr
15	Thr	Leu 370	Phe	Asn	Lys	Ile	Tyr 375	Arg	Arg	Ala	Phe	Ser 380	Asn	Tyr	Leu	Arg
	Cys 385	Asn	Tyr	Lys	Val	Glu 390	Lys	Lys	Pro	Pro	Val 395	Arg	Gln	Ile	Pro	Arg 400
	Val	Ala	Ala	Thr	Ala 405	Leu	Ser	Gly	Arg	Glu 410	Leu	Asn	Val	Asn	Ile 415	Tyr
20	Arg	His	Thr	Asn 420	Glu	Pro	Val	Ile	Glu 425	Lys	Ala	Ser	Asp	Asn 430	Glu	Pro
	Gly	Ile	Glu 435	Met	Gln	Val	Glu	Asn 440	Leu	Glu	Leu	Pro	Val 445	Asn	Pro	Ser
25	Ser	Val 450	Val	Ser	Glu	Arg	Ile 455	Ser	Ser	Val						
	(232) IN	FORM	OITA	1 FOI	SEÇ	O ID	NO:2	231:						,		
	(i)	(A)	LEI	CE CH NGTH: PE: r	106	8 ba	ase p		3							
30		(C)	STI	RANDE	EDNES	SS: 8	singl	.e								,
	(i:	i) M(DLECT	JLE 7	YPE:	: DN?	4 (ge	nomi	lc)		•			÷		
	(x :	i) SI	EQUEI	NCE I	ESCF	RIPT	ON:	SEQ	ID N	10:23	31:					
	ATGGATCA	GT TO	CCT	GAAT	AG1	rgac?	AGAA	AACI	TTG	GT A	CGAT	rgat 1	T GO	GCTG#	AGGCC	: 60
35	TGTTATAT	rg go	GAC)	ATCG1	GG1	CTT	rggg	ACTO	TGT	rcc 1	GTC	CATA	T CI	TACTO	CCGTC	120
	ልጥርጥ ተጥርር	יידי מיי	וממרנ	ጉጥርር፣	י מממ	ימממב	ביתיתים	ጥጥርያ	יייאכיי	וכיתי יו	ጥርሮ	ירייריז	יר כז	\ አ ሮልር	ברא א כ	2 100

PCT/US99/23938

				199			
	AAGCCCAAGA	GTGTCACCGA	CATTTACCTC	CTGAACCTGG	CCTTGTCTGA	TCTGCTGTTT	240
	GTAGCCACTT	TGCCCTTCTG	GACTCACTAT	TTGATAAATG	AAAAGGGCCT	CCACAATGCC	300
	ATGTGCAAAT	TCACTACCGC	CTTCTTCTTC	ATCGGCTTTT	TTGGAAGCAT	ATTCTTCATC	360
	ACCGTCATCA	GCATTGATAG	GTACCTGGCC	ATCGTCCTGG	CCGCCAACTC	CATGAACAAC	420
5	CGGACCGTGC	AGCATGGCGT	CACCATCAGC	CTAGGCGTCT	GGGCAGCAGC	CATTTTGGTG	480
	GCAGCACCCC	AGTTCATGTT	CACAAAGCAG	AAAGAAAATG	AATGCCTTGG	TGACTACCCC	540
	GAGGTCCTCC	AGGAAATCTG	GCCCGTGCTC	CGCAATGTGG	AAACAAATTT	TCTTGGCTTC	600
	CTACTCCCCC	TGCTCATTAT	GAGTTATTGC	TACTTCAGAA	TCATCCAGAC	GCTGTTTTCC	660
	TGCAAGAACC	ACAAGAAAGC	CAAAGCCAAG	AAACTGATCC	TTCTGGTGGT	CATCGTGTTT	720
10	TTCCTCTTCT	GGACACCCTA	CAACGTTATG	ATTTTCCTGG	AGACGCTTAA	GCTCTATGAC	780
	TTCTTTCCCA	GTTGTGACAT	GAGGAAGGAT	CTGAGGCTGG	CCCTCAGTGT	GACTGAGACG	840
	GTTGCATTTA	GCCATTGTTG	CCTGAATCCT	CTCATCTATG	CATTTGCTGG	GGAGAAGTTC	900
	AGAAGATACC	TTTACCACCT	GTATGGGAAA	TGCCTGGCTG	TCCTGTGTGG	GCGCTCAGTC	960
	CACGTTGATT	TCTCCTCATC	TGAATCACAA	AGGAGCAGGC	ATGGAAGTGT	TCTGAGCAGC	1020
15	AATTTTACTT	ACCACACGAG	TGATGGAGAT	GCATTGCTCC	TTCTCTGA		1068
•	(233) INFO	RMATION FOR	SEQ ID NO:	232:			
20		SEQUENCE CH (A) LENGTH: (B) TYPE: ai (C) STRANDEI (D) TOPOLOG	355 amino a mino acid DNESS: Y: not rele	acids			
	(11)	MOLECULE T	rrm: protel	LI			

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:232:

Met Asp Gln Phe Pro Glu Ser Val Thr Glu Asn Phe Glu Tyr Asp Asp 15

Leu Ala Glu Ala Cys Tyr Ile Gly Asp 25

Phe Leu Ser Ile Phe Tyr Ser Val Ile Phe Ala Ile Gly Leu Val Gly 45

Asn Leu Val Val Val Phe Ala Leu Thr Asn Ser Lys Lys Pro Lys Ser Ser 55

Val Thr Asp Ile Tyr Leu Leu Asn Leu Ala Leu Ser Asp Leu Leu Phe

	65					70					75					80
	Val	Ala	Thr	Leu	Pro 85	Phe	Trp	Thr	His	Tyr 90	Leu	Ile	Asn	Glu	Lys 95	Gly
5	Leu	. His	Asn	Ala 100	Met	Cys	Lys	Phe	Thr 105		Ala	Phe	Phe	Phe		Gly
	Phe	Phe	Gly 115	Ser	Ile	Phe	Phe	Ile 120		Val	Ile	Ser	Ile 125	Asp	Arg	Tyr
	Leu	Ala 130	Ile	Val	Leu	Ala	Ala 135		Ser	Met	Asn	Asn 140	Arg	Thr	Val	Gln
10	His 145	Gly	Val	Thr	Ile	Ser 150	Leu	Gly	Val	Trp	Ala 155	Ala	Ala	Ile	Leu	Val 160
	Ala	Ala	Pro	Gln	Phe 165	Met	Phe	Thr	Lys	Gln 170	Lys	Glu	Asn	Glu	Cys 175	Leu
15	Gly	Asp	Tyr	Pro 180	Glu	Val	Leu	Gln	Glu 185	Ile	Trp	Pro	Val	Leu 190	Arg	Asn
	Val	Glu	Thr 195	Asn	Phe	Leu	Gly	Phe 200	Leu	Leu	Pro	Leu	Leu 205	Ile	Met	Ser
	Tyr	Cys 210	Tyr	Phe	Arg	Ile	Ile 215	Gln	Thr	Leu	Phe	Ser 220	Cys	Lys	Asn	His
20	Lys 225	Lys	Ala	Lys	Ala	Lys 230	Lys	Leu	Ile	Leu	Leu 235	Val	Val	Ile	Val	Phe 240
	Phe	Leu	Phe	Trp	Thr 245,	Pro	Tyr	Asn	Val	Met 250	Ile	Phe	Leu	Glu	Thr 255	Leu
25	Lys	Leu	Tyr	Asp 260	Phe	Phe	Pro	Ser	Cys 265	Asp	Met	Arg	Lys	Asp 270	Leu	Arg
	Leu	Ala	Leu 275	Ser	Val	Thr	Glu	Thr 280	Val	Ala	Phe	Ser	His 285	Cys	Cys	Leu
	Asn	Pro 290	Leu	Ile	Tyr		Phe 295	Ala	Gly	Glu	Lys	Phe 300	Arg	Arg	Tyr	Leu
30	Tyr 305	His	Leu	Tyr	Gly	Lys 310	Cys	Leu	Ala	Val	Leu 315	Cys	Gly	Arg		Val 320
	His	Val	Asp	Phe	Ser 325	Ser	Ser	Glu		Gln 330	Arg	Ser	Arg		Gly 335	Ser
35	Val	Leu	Ser	Ser 340	Asn	Phe	Thr	Tyr	His 345	Thr	Ser	Asp		Asp 350	Ala	Leu
	Leu	Leu	Leu 355													

בי באסריביתה האות הויתחפותם

WO 00/22129 PCT/US99/23938

	(234) INFORMATION FOR SEQ ID NO:233:	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 29 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(iv) ANTI-SENSE: NO	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:233:	
10	GGCTTAAGAG CATCATCGTG GTGCTGGTG	29
	(235) INFORMATION FOR SEQ ID NO:234:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 34 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(iv) ANTI-SENSE: YES	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:234:	
20	GTCACCACCA GCACCACGAT GATGCTCTTA AGCC	34
	(236) INFORMATION FOR SEQ ID NO:235:	
25	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	. •
	(ii) MOLECULE TYPE: DNA (genomic) (xi) SEQUENCE DESCRIPTION: SEQ ID NO:235:	
	CAAAGAAAGT ACTGGGCATC GTCTTCTTCC T	31
30	(237) INFORMATION FOR SEQ ID NO:236:	
	(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 30 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single	
35	(D) TOPOLOGY: linear	•
	(ii) MOLECULE TYPE: DNA (genomic)	

	(X1) SEQUENCE DESCRIPTION: SEQ ID NO:236:	
	TGCTCTAGAT TCCAGATAGG TGAAAACTTG	30
	(238) INFORMATION FOR SEQ ID NO.237:	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 50 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
10	(iv) ANTI-SENSE: NO	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:237:	
	CTAGGGGCAC CATGCAGGCT ATCAACAATG AAAGAAAAGC TAAGAAAGTC	50
	(239) INFORMATION FOR SEQ ID NO:238:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 50 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
20	(iv) ANTI-SENSE: YES	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:238:	
	CAAGGACTTT CTTAGCTTTT CTTTCATTGT TGATAGCCTG CATGGTGCCC	50
	(240) INFORMATION FOR SEQ ID NO:239:	
25	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 35 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
30	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:239:	
	CGGCGGCAGA AGGCGAAACG CATGATCCTC GCGGT	35
	(241) INFORMATION FOR SEQ ID NO:240:	
35	(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 35 base pairs(B) TYPE: nucleic acid	

PCT/US99/23938

		(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA (genomic)	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO:240:	
5	ACCGC	GAGGA TCATGCGTTT CGCCTTCTGC CGCCG	35
	(242)	INFORMATION FOR SEQ ID NO:241:	
10		(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA (genomic)	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO:241:	
	GAGAC	ATATT ATCTGCCACG GAGG	24
15	(243)	INFORMATION FOR SEQ ID NO:242:	
20		 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
		(ii) MOLECULE TYPE: DNA (genomic)	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO:242:	
	TTGGC	ATAGA AACCGGACCC AAGG	24
	(244)	INFORMATION FOR SEQ ID NO:243:	
25		 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 28 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
30		(ii) MOLECULE TYPE: DNA (genomic)	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO:243:	
	TAAGA	ATTCC ATAAAATTA TGGAATGG	28
	(245)	INFORMATION FOR SEQ ID NO:244:	
		(i) SEQUENCE CHARACTERISTICS:	

	(A) LENGTH: 30 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
5	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:244:	
	CCAGGATCCA GCTGAAGTCT TCCATCATTC	30
	(246) INFORMATION FOR SEQ ID NO:245:	
10	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1071 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
15	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:245:	
	ATGAATGGGG TCTCGGAGGG GACCAGAGGC TGCAGTGACA GGCAACCTGG GGTCCTGACA	60
	CGTGATCGCT CTTGTTCCAG GAAGATGAAC TCTTCCGGAT GCCTGTCTGA GGAGGTGGGG	120
	TCCCTCCGCC CACTGACTGT GGTTATCCTG TCTGCGTCCA TTGTCGTCGG AGTGCTGGGC	180
•	AATGGGCTGG TGCTGTGGAT GACTGTCTTC CGTATGGCAC GCACGGTCTC CACCGTCTGC	240
20,	TTCTTCCACC TGGCCCTTGC CGATTTCATG CTCTCACTGT CTCTGCCCAT TGCCATGTAC	300
	TATATTGTCT CCAGGCAGTG GCTCCTCGGA GAGTGGGCCT GCAAACTCTA CATCACCTTT	360
	GTGTTCCTCA GCTACTTTGC CAGTAACTGC CTCCTTGTCT TCATCTCTGT GGACCGTTGC	420
*	ATCTCTGTCC TCTACCCCGT CTGGGCCCTG AACCACCGCA CTGTGCAGCG GGCGAGCTGG	480
	CTGGCCTTTG GGGTGTGGCT CCTGGCCGCC GCCTTGTGCT CTGCGCACCT GAAATTCCGG	540
25	ACAACCAGAA AATGGAATGG CTGTACGCAC TGCTACTTGG CGTTCAACTC TGACAATGAG	600
	ACTGCCCAGA TTTGGATTGA AGGGGTCGTG GAGGGACACA TTATAGGGAC CATTGGCCAC	660
	TTCCTGCTGG GCTTCCTGGG GCCCTTAGCA ATCATAGGCA CCTGCGCCCA CCTCATCCGG	720
	GCCAAGCTCT TGCGGGAGGG CTGGGTCCAT GCCAACCGGC CCGCGAGGCT GCTGCTGGTG	780
	CTGGTGAGCG CTTTCTTTAT CTTCTGGTCC CCGTTTAACG TGGTGCTGTT GGTCCATCTG	840
30	TGGCGACGGG TGATGCTCAA GGAAATCTAC CACCCCGGA TGCTGCTCAT CCTCCAGGCT	900
	AGCTTTGCCT TGGGCTGTGT CAACAGCAGC CTCAACCCCT TCCTCTACGT CTTCGTTGGC	960

BRIGHOUSE SINO CONSTROATE

	AGAGATTT	CC A	AGAA	aagt	T TT	TCCA	GTCT	TTG	ACTT	CTG ·	CCCT	GGCG	AG G	GCGT	TTGG.	A 10	20
	GAGGAGGA	GT T	TCTG	TCAT	C CT	GTCC	CCGT	GGC.	AACG	ccc ·	CCCG	GGAA	TG A			10	71
	(247) IN	FORM	ATIO	N FO	R SE	Q ID	NO:	246:									
5	(i	(B)	QUENC) LEI) TYI) STI) TOI	ngth PE: a RANDI	: 350 amino EDNE:	6 am 5 ac: SS:	ino a	acid	s ··								
	(i:	i) M(OLEC	JLE :	TYPE	: pro	otei	n									
10	(x:	i) SI	EQUE	NCE 1	DESCI	RIPT	ION:	SEQ	ID 1	NO : 2	46:						
	Met 1	Asn	Gly	Val	Ser 5	Glu	Gly	Thr	Arg	Gly 10	Cys	Ser	Asp	Arg	Gln 15	Pro	
	Gly	Val	Leu	Thr 20	Arg	Asp	Arg	Ser	Cys 25	Ser	Arg	Lys	Met	Asn 30	Ser	Ser	**
15	Gly	Cys	Leu 35	Ser	Glu	Glu	Val	Gly 40	Ser	Leu	Arg	Pro	Leu 45	Thr	Val	Val	*.
-	Ile	Leu 50	Ser	Ala	Ser	Ile	Val 55	Val	Gly	Val	Leu	Gly 60	Asn	Gly	Leu	Val	
20	Leu 65	Trp	Met	Thr	Val	Phe 70	Arg	Met	Ala		Thr 75	Val	Ser	Thr	Val	Cys 80	•
	Phe	Phe	His	Leu	Ala 85	Leu	Ala	Asp	Phe	Met 90	Leu	Ser	Leu	Ser	Leu 95	Pro	्रम् स्रोत
	Ile	Ala	Met	Tyr 100	Tyr	Ile	Val	Ser	Arg 105	Gln	Trp	Leu	Leu	Gly 110	Glu	Trp	
25	Ala	Cys	Lys 115	Leu	Tyr	Ile	Thr	Phe 120	Val	Phe	Leu	Ser	Tyr 125	Phe	Ala	Ser	
	Asn	Cys 130	Leu	Leu	Val	Phe	Ile 135	Ser	Val	Asp	Arg	Cys 140	Ile	Ser	Val	Leu	
30	Tyr 145	Pro	Val	Trp	Ala	Leu 150	Asn	His	Arg	Thr	Val 155	Gln	Arg	Ala	Ser	Trp 160	
	Leu	Ala	Phe	Gly	Val 165	Trp	Leu	Leu	Ala	Ala 170	Ala	Leu	Cys	Ser	Ala 175	His	
	Leu	Lys	Phe	Arg 180	Thr	Thr	Arg	Lys	Trp 185	Asn	Gly	Cys	Thr	His 190	Cys	Tyr	
35	Leu	Ala	Phe 195	Asn	Ser	Asp	Asn	Glu 200	Thr	Ala	Gln	Ile	Trp 205	Ile	Glu	Gly	

	Val	Val 210	Glu	Gly	His	Ile	Ile 215	Gly	Thr	Ile	Gly	His 220	Phe	Leu	Leu	Gly	
	Phe 225	Leu	Gly	Pro	Leu	Ala 230	Ile	Ile	Gly	Thr	Cys 235	Ala	His	Leu	Ile	Arg 240	
5	Ala	Lys	Leu	Leu	Arg 245	Glu	Gly	Trp	Val	His 250	Ala	Asn	Arg	Pro	Ala 255	Arg	
	Leu	Leu	Leu	Val 260	Leu	Val	Ser	Ala	Phe 265	Phe	Ile	Phe	Trp	Ser 270	Pro	Phe	
10	Asn		Val 275	Leu	Leu	Val	His	Leu 280	Trp	Arg	Arg	Val	Met 285	Leu	Lys	Glu	
	Ile	Tyr 290	His	Pro	Arg	Met	Leu 295	Leu	Ile	Leu	Gln	Ala 300	Ser	Phe	Ala	Leu	
	Gly 305	Cys	Val	Asņ	Ser	Ser 310	Leu	Asn	Pro	Phe	Leu 315	Tyr	Val	Phe	Val	Gly 320	
15	Arg	Asp	Phe	Gln	Glu 325	Lys	Phe	Phe	Gln	Ser 330	Leu	Thr	Ser	Ala	Leu 335	Ala	
	Arg	Ala	Phe	Gly 340	Glu	Glu	Glu	Phe	Leu 345	Ser	Ser	Cys	Pro	Arg 350	Gly	Asn	
20	Ala		Arg 355	Glu													
	(248) INF	ORMA	TION	FOF	SEC) ID	NO:2	47:					•				
25	(i)	(B)	LEN TYP STR	IGTH: PE: n PANDE	32 ucle DNES	base	e pai acid singl	.rs			- ,						
	(ii) MO	LECU	LE I	YPE:	DNA	4 (ge	nomi	.c)								
	(xi) SE	QUEN	ICE I	ESCF	e i PTI	ON:	SEQ	ID N	IO:24	7:		٠				
	GCAGAATTC	G GC	GGCC	CCAI	GGA	CCTG	SCCC	CC									32
30	(249) INF	ORMA	TION	FOR	SEC) ID	NO:2	48:									
35	(i)	(B) (C)	LEN TYP STR	IGTH: PE: II PANDE	30 ucle DNES	base	e pai scid singl	rs.									
	(ii) MO	LECU	LE T	YPE:	DNA	A (ge	nomi	.c)								
	(xi) SE	QUEN	ICE D	ESCF	IPTI	ON:	SEQ	ID N	10:24	8:						

WO 00/22129 PCT/US99/23938

207

GCTGGATCCC CCGAGCAGTG GCGTTACTTC 30													
(250) INFORMATION FOR SEQ ID NO:249:													
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 903 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 													
(ii) MOLECULE TYPE: DNA (genomic)													
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:249:													
ATGGACCTGC CCCCGCAGCT CTCCTTCGGC CTCTATGTGG CCGCCTTTGC GCTGGGCTTC	60												
CCGCTCAACG TCCTGGCCAT CCGAGGCGCG ACGGCCCACG CCCGGCTCCG TCTCACCCCT	120												
AGCCTGGTCT ACGCCCTGAA CCTGGGCTGC TCCGACCTGC TGCTGACAGT CTCTCTGCCC	180												
CTGAAGGCGG TGGAGGCGCT AGCCTCCGGG GCCTGGCCTC TGCCGGCCTC GCTGTGCCCC	240												
GTCTTCGCGG TGGCCCACTT CTTCCCACTC TATGCCGGCG GGGGCTTCCT GGCCGCCCTG	300												
AGTGCAGGCC GCTACCTGGG AGCAGCCTTC CCCTTGGGCT ACCAAGCCTT CCGGAGGCCG	360												
TGCTATTCCT GGGGGGTGTG CGCGGCCATC TGGGCCCTCG TCCTGTGTCA CCTGGGTCTG	420												
GTCTTTGGGT_TGGAGGCTCC AGGAGGCTGG CTGGACCACA GCAACACCTC CCTGGGCATC	480												
AACACACCGG TCAACGGCTC TCCGGTCTGC CTGGAGGCCT GGGACCCGGC CTCTGCCGGC	540												
CCGGCCCGCT TCAGCCTCTC TCTCCTGCTC TTTTTTCTGC CCTTGGCCAT CACAGCCTTC	600												
TGCTACGTGG GCTGCCTCCG GGCACTGGCC CGCTCCGGCC TGACGCACAG GCGGAAGCTG	660												
CGGGCCGCCT GGGTGGCCGG CGGGGCCCTC CTCACGCTGC TGCTCTGCGT AGGACCCTAC	720												
AACGCCTCCA ACGTGGCCAG CTTCCTGTAC CCCAATCTAG GAGGCTCCTG GCGGAAGCTG	780												
GGGCTCATCA CGGGTGCCTG GAGTGTGGTG CTTAATCCGC TGGTGACCGG TTACTTGGGA	840												
AGGGGTCCTG GCCTGAAGAC AGTGTGTGCG GCAAGAACGC AAGGGGGCAA GTCCCAGAAG	900												
TAA	903												
(251) INFORMATION FOR SEQ ID NO:250:													

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 300 amino acids
 - (B) TYPE: amino acid
- (C) STRANDEDNESS:

5

10

15

20

25

30

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein(xi) SEQUENCE DESCRIPTION: SEQ ID NO:250:																	
		Met 1	Asp	Leu	Pro	Pro	Gln	Leu	Ser	Phe	Gly	Leu	Туг	Val	Ala	Ala 15	Phe
5		Ala	Leu	Gly	Phe 20	Pro	Leu	Asn	Val	Leu 25	Ala	Ile	Arg	Gly	Ala 30	Thr	Ala
		His	Ala	Arg 35	Leu	Arg	Leu	Thr	Pro	Ser	Leu	Val	Tyr	Ala 45	Leu	Asn	Leu
10		Gly	Cys 50	Ser	Asp	Leu	Leu	Leu 55	Thr	Val	Ser	Leu	Pro 60	Leu	Lys	Ala	Val
,		Glu 65	Ala	Leu	Ala	Ser	Gly 70	Ala	Trp	Pro	Leu	Pro 75	Ala	Ser	Leu	Cys	Pro 80
		Val	Phe	Ala	Val	Ala 85	His	Phe	Phe	Pro	Leu 90	Tyr	Ala	Gly	Gly	Gly 95	Phe
15		Leu	Ala	Ala	Leu 100	Ser	Ala	Gly	Arg	Tyr 105	Leu	Gly	Ala	Ala	Phe 110	Pro	Leu
	•	Gly	Tyr	Gln 115	Ala	Phe	Arg	Arg	Pro 120	Cys	Tyr	Ser	Trp	Gly 125	Val	Cys	Ala
20		Ala	Ile 130	Trp	Ala	Leu	Val	Leu 135	Cys	His	Leu	Gly	Leu 140	Val	Phe	Gly	Leu
		Glu 145	Ala	Pro	Gly	Gly	Trp 150	Leu	Asp	His	Ser	Asn 155	Thr	Ser	Leu	Gly	Ile 160
		Asn	Thr	Pro	Val	Asn 165	Gly	Ser	Pro	Val	Cys 170	Leu	Glu	Ala	Trp	Asp 175	Pro
25		Ala	Ser	Ala	Gly 180	Pro	Ala	Arg	Phe	Ser 185	Leu	Ser	Leu	Leu	Leu 190	Phe	Phe
		Leu	Pro	Leu 195	Ala	Ile	Thr	Ala	Phe 200	Cys	Tyr	Val	Gly	Cys 205	Leu	Arg	Ala
30		Leu	Ala 210	Arg	Ser	Gly	Leu	Thr 215	His	Arg	Arg	Lys	Leu 220	Arg	Ala	Ala	Trp
		Val 225	Ala	Gly	Gly	Ala	Leu 230	Leu	Thr	Leu	Leu	Leu 235	Cys	Val	Gly-	Pro	Tyr 240
		Asn	Ala	Ser	Asn	Val 245	Ala	Ser	Phe	Leu	Tyr 250	Pro	Asn	Leu	Gly	Gly 255	Ser
35		Trp	Arg	Lys	Leu 260	Gly	Leu	Ile	Thr	Gly 265	Ala	Trp	Ser		Val 270	Leu	Asn
		Pro	Leu	Val	Thr	Gly	Tyr	Leu	Gly	Arg	Gly	Pro	Gly	Leu	Lys	Thr	Val

WO 00/22129 PCT/US99/23938

	275 280 285	
	Cys Ala Ala Arg Thr Gln Gly Gly Lys Ser Gln Lys 290 295 300	
	(252) INFORMATION FOR SEQ ID NO:251:	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
10	(ii) MOLECULE TYPE: DNA (genomic)	-
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:251:	
	CTCAAGCTTA CTCTCTCA CCAGTGGCCA C	31
	(253) INFORMATION FOR SEQ ID NO:252:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	•
	(ii) MOLECULE TYPE: DNA (genomic)	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:252:	
	CCCTCCTCCC CCGGAGGACC TAGC	24
	(254) INFORMATION FOR SEQ ID NO:253:	
25	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1041 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
•	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:253:	
30	ATGGATACAG GCCCCGACCA GTCCTACTTC TCCGGCAATC ACTGGTTCGT CTTCTCGGTG	60
	TACCTTCTCA CTTTCCTGGT GGGGCTCCCC CTCAACCTGC TGGCCCTGGT GGTCTTCGTG	120
	GGCAAGCTGC AGCGCCGCCC GGTGGCCGTG GACGTGCTCC TGCTCAACCT GACCGCCTCG	180
	GACCTGCTCC TGCTGCTGTT CCTGCCTTTC CGCATGGTGG AGGCAGCCAA TGGCATGCAC	240
	TGGCCCCTGC CCTTCATCCT CTGCCCACTC TCTGGATTCA TCTTCTTCAC CACCATCTAT	300

	CTCACCGCCC	TCTTCCTGGC	AGCTGTGAGC	ATTGAACGCT	TCCTGAGTGT	GGCCCACCCA	360
	CTGTGGTACA	AGACCCGGCC	GAGGCTGGGG	CAGGCAGGTC	TGGTGAGTGT	GGCCTGCTGG	420
	CTGTTGGCCT	CTGCTCACTG	CAGCGTGGTC	TACGTCATAG	AATTCTCAGG	GGACATCTCC	480
	CACAGCCAGG	GCACCAATGG	GACCTGCTAC	CTGGAGTTCC	GGAAGGACCA	GCTAGCCATC	540
5	CTCCTGCCCG	TGCGGCTGGA	GATGGCTGTG	GTCCTCTTTG	TGGTCCCGCT	GATCATCACC	600
	AGCTACTGCT	ACAGCCGCCT	GGTGTGGATC	CTCGGCAGAG	GGGGCAGCCA	CCGCCGGCAG	660
	AGGAGGGTGG	CGGGGCTGTT	GGCGGCCACG	CTGCTCAACT	TCCTTGTCTG	CTTTGGGCCC	720
	TACAACGTGT	CCCATGTCGT	GGGCTATATC	TGCGGTGAAA	GCCCGGCATG	GAGGATCTAC	780
	GTGACGCTTC	TCAGCACCCT	GAACTCCTGT	GTCGACCCCT	TTGTCTACTA	CTTCTCCTCC	840
10	TCCGGGTTCC	AAGCCGACTT	TCATGAGCTG	CTGAGGAGGT	TGTGTGGGCT	CTGGGGCCAG	900
	TGGCAGCAGG	AGAGCAGCAT	GGAGCTGAAG	GAGCAGAAGG	GAGGGGAGGA	GCAGAGAGCG	960
•	GACCGACCAG	CTGAAAGAAA	GACCAGTGAA	CACTCACAGG	GCTGTGGAAC	TGGTGGCCAG	1020
	GTGGCCTGTG	CTGAAAGCTA	G				1041
•							

(255) INFORMATION FOR SEQ ID NO:254:

15 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 346 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- 20 (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:254:

Met Asp Thr Gly Pro Asp Gln Ser Tyr Phe Ser Gly Asn His Trp Phe 1 5 10 15

Val Phe Ser Val Tyr Leu Leu Thr Phe Leu Val Gly Leu Pro Leu Asn 25 25 30

Leu Leu Ala Leu Val Val Phe Val Gly Lys Leu Gln Arg Arg Pro Val 35 40 45

Ala Val Asp Val Leu Leu Leu Asn Leu Thr Ala Ser Asp Leu Leu Leu 50 55 60

Leu Leu Phe Leu Pro Phe Arg Met Val Glu Ala Ala Asn Gly Met His

65 70 75 80

Trp Pro Leu Pro Phe Ile Leu Cys Pro Leu Ser Gly Phe Ile Phe Phe 85 90 95

		Thr	Thr	Ile	Tyr 100	Leu	Thr	Ala	Leu	Phe 105	Leu	Ala	Ala	Val	Ser 110	Ile	Glu
		Arg	Phe	Leu 115	Ser	Val	Ala	His	Pro 12C	Leu	Trp	Tyr	Lys	Thr 125	Arg	Pro	Arg
5		Leu	Gly 130	Gln	Ala	Gly	Leu	Val 135	Ser	Val	Ala	Cys	Trp 140	Leu	Leu	Ala	Ser
		Ala 145	His	Cys	Ser	Val	Val 150	Tyr	Val	Ile	Glu	Phe 155	Ser	Gly	Asp	Ile	Ser 160
10	:	His	Ser	Gln	Gly	Thr 165	Asn	Gly	Thr	Cys	Tyr 170	Leu	Glu	Phe	Arg	Lys 175	Asp
	•	Gln	Leu	Ala	Ile 180	Leu	Leu	Pro	Val	Arg 185	Leu	Glu	Met	Ala	Val 190	Val	Leu
	:	Phe	Val	Val 195	Pro	Leu	Ile	Ile	Thr 200	Ser	Tyr	Cys	Tyr	Ser 205	Arg	Leu	Val
15	•	Trp	Ile 210	Leu	Gly	Arg	Gly	Gly 215	Ser	His	Arg	Arg	Gln 220	Arg	Arg	Val	Ala
		Gly 225	Leu	Leu	Ala	Ala	Thr 230	Leu	Leu	Asn	Phe	Leu 235	Val	Cys	Phe	Gly	Pro 240
20	•	Tyr	Asn	Val	Ser	His 245	Val	Val	Gly		Ile 250	Cys	Gly	Glu	Şer	Pro 255	Ala
	•	Trp	Arg	Ile	Туг 260	Val	Thr	Leu	Leu	Ser 265	Thr	Leu	Asn	Ser	Cys 270	Val	Asp
	1	Pro		Val 275	Tyr	Tyr	Phe	Ser	Ser 280	Ser	Gly	Phe	Gln	Ala 285	Asp	Phe	His
25			Leu 290	Leu	Arg	Arg	Leu	Cys 295	Gly	Leu	Trp	Gly	Gln 300	Trp	Gln	Gln	Glu
		Ser 305	Ser	Met	Glu	Leu	Lys 310	Glu	Gln	Lys	Gly	Gly 315	Glu	Glu	Gln	Arg	Ala 320
30	1	Asp	Arg	Pro	Ala	Glu 325	Arg	Lys	Thr	Ser	Glu 330	His	Ser	Gln	Gly	Cys 335	Gly
	7	Thr	Gly	Gly	Gln 340	Val	Ala	Cys	Ala	Glu 345	Ser						
	(256)	INF	ORMA	TION	FOR	SEO	ID	NO:2	55:								

35

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:255:	
	TTTAAGCTTC CCCTCCAGGA TGCTGCCGGA C	31
	(257) INFORMATION FOR SEQ ID NO:256:	
5	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: not relevant	
10	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:256:	
	GGCGAATTCT GAAGGTCCAG GGAAACTGCT A	31
	(258) INFORMATION FOR SEQ ID NO:257:	
15	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 993 base pairs	
:	(B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
:	(ii) MOLECULE TYPE: DNA (genomic)	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:257:	
	ATGCTGCCGG ACTGGAAGAG CTCCTTGATC CTCATGGCTT ACATCATCAT CTTCCTCACT	60
	GGCCTCCCTG CCAACCTCCT GGCCCTGCGG GCCTTTGTGG GGCGGATCCG CCAGCCCCAG	120
	CCTGCACCTG TGCACATCCT CCTGCTGAGC CTGACGCTGG CCGACCTCCT CCTGCTGCTG	180
	CTGCTGCCCT TCAAGATCAT CGAGGCTGCG TCGAACTTCC GCTGGTACCT GCCCAAGGTC	240
25	GTCTGCGCCC TCACGAGTTT TGGCTTCTAC AGCAGCATCT ACTGCAGCAC GTGGCTCCTG	300
	GCGGGCATCA GCATCGAGCG CTACCTGGGA GTGGCTTTCC CCGTGCAGTA CAAGCTCTCC	360
	CGCCGGCCTC TGTATGGAGT GATTGCAGCT CTGGTGGCCT GGGTTATGTC CTTTGGTCAC	420
	TGCACCATCG TGATCATCGT TCAATACTTG AACACGACTG AGCAGGTCAG AAGTGGCAAT	480
	GAAATTACCT GCTACGAGAA CTTCACCGAT AACCAGTTGG ACGTGGTGCT GCCCGTGCGG	540
30	CTGGAGCTGT GCCTGGTGCT CTTCTTCATC CCCATGGCAG TCACCATCTT CTGCTACTGG	600
	CGTTTTGTGT GGATCATGCT CTCCCAGCCC CTTGTGGGGG CCCAGAGGCG GCGCCGAGCC	660
	GTGGGGCTGG CTGTGGTGAC GCTGCTCAAT TTCCTGGTGT GCTTCGGACC TTACAACGTG	720

	TCCCACCI	rgg I	GGGG	TATO	CA CC	AGAG	AAAA	AGC	CCCI	GGT	GGCG	GTC#	AT A	AGCCC	TGGT	G	780
	TTCAGTTC	CAC I	CAAC	GCCA	G TC	TGGA	cccc	CTG	CTCI	TCT	ATTI	CTCI	TC 1	TCAC	TGGT	'G	840
	CGCAGGGC	AT I	TGGG	AGAG	G GC	TGCA	GGTG	CTG	CGGA	ATC	AGGG	CTCC	TC C	CTGI	TGGG	A	900
	CGCAGAGG	CA A	AGAC	ACAG	C AG	AGGG	GACA	AAT	GAGG	ACA	GGGG	TGTG	GG 1	CAAG	GAGA	A	960
5	GGGATGCC	AA G	TTCG	GACT	T CA	CTAC	AGAG	TAG	}								993
	(259) IN	FORM	OITA	N FO	R SE	Q ID	NO:	258:									
10		(B (C) LE) TY) ST) TO	NGTH PE: RAND POLO	: 36 amin EDNE GY:	2 am o ac SS: not	ino id rele	acid vant									
		i) S				_			ID	NO:2	58:						
												Met	Ala	Tyr	Ile	Ile	41
15	1				5					10					15		
	Ile	Phe	Leu	Thr 20	Gly	Leu	Pro	Ala	Asn 25	Leu	Leu	Ala	Leu	Arg 30	Ala	Phe	
	Val	Gly	Arg 35	Ile	Arg	Gln	Pro	Gln 40	Pro	Ala	Pro	Val	His 45	Ile	Leu	Leu	
20	Leu	Ser 50	Leu	Thr	Leu	Ala	Asp 55	Leu	Leu	Leu	Leu	Leu 60	Leu	Leu	Pro	Phe	
	Lys 65	Ile	Ile	Glu	Ala	Ala 70	Ser	Asn	Phe	Arg	Trp 75	Tyr	Leu	Pro	Lys	Val 80	
25	Val	Cys	Ala	Leu	Thr 85	Ser	Phe	Gly	Phe	Tyr 90	Ser	Ser	Ile	Tyr	Суs 95	Ser	
	Thr	Trp	Leu	Leu 100	Ala	Gly	Ile	Ser	Ile 105	Glu	Arg	Tyr	Leu	Gly 110	Val	Ala	
	Phe	Pro	Val 115	Gln	Tyr	Lys	Leu	Ser 120	Arg	Arg	Pro	Leu	Tyr 125	Gly	Val	Ile	
30	Ala	Ala 130	Leu	Val	Ala	Trp	Val 135	Met	Ser	Phe	Gly	His 140	Cys	Thr	Ile	Val	
	Ile 145	Ile	Val	Gln	Tyr	Leu 150	Asn	Thr	Thr	Glu	Gln 155	Val	Arg	Ser	Gly	Asn 160	
35	Glu	Ile	Thr	Cys	Tyr 165	Glu	Asn	Phe	Thr	Asp 170	Asn	Gln	Leu	Asp	Val 175	Val	

		Leu	Pro	Val	Arg 180	Leu	Glu	Leu	Cys	Leu 185	Val	Leu	Phe	Phe	Ile 190	Pro	Met
		Ala	Val	Thr 195	Ile	Phe	Cys	Tyr	Trp 200	Arg	Phe	Val	Trp	Ile 205	Met	Leu	Ser
5		Gln	Pro 210	Leu	Val	Gly	Ala	Gln 215	Arg	Arg	Arg	Arg	Ala 220	Val	Gly	Leu	Ala
		Val 225	Val	Thr	Leu	Leu	Asn 230	Phe	Leu	Val	Cys	Phe 235	Gly	Pro	Tyr	Asn	Val 240
10		Ser	His	Leu	Val	Gly 245	Tyr	His	Gln	Arg	Lys 250	Ser	Pro	Trp	Trp	Arg 255	Ser
		Ile	Ala	Val	Val 260	Phe	Ser	Ser	Leu	Asn 265	Ala	Ser	Leu	Asp	Pro 270	Leu	Leu
		Phe	Tyr	Phe 275	Ser	Ser	Ser	Val	Val 280	Arg	Arg	Ala	Phe	Gly 285	Arg	Gly	Leu
15		Gln	Val 290	Leu	Arg	Asn	Gln	Gly 295	Ser	Ser	Leu	Leu	Gly 300	Arg	Arg	Gly	Lys
		Asp 305	Thr	Ala	Glu	Gly	Thr 310	Asn	Glu	Asp		Gly 315	Val	Gly	Gln	_	Glu 320
20		Gly	Met	Pro	Ser	Ser 325	Asp	Phe	Thr	Thr	Glu 330						
	(260)	INF	ORMA	TION	FOR	SEQ	ID	NO:2	59:		,						
25		(i)	(A) (B) (C)	LEN TYP STR	GTH: E: n ANDE	30 ucle DNES	base ic a S: s	pai cid ingl	rs								
		(ii		TOP					nomi	۵۱		•					
									SEQ		0:25	9:					
	CCCAA																30
30	(261)			TION		SEQ	ID :	NO : 2	60:								
35	s •••	(i)	(A) (B) (C)	UENC LENG TYP: STR.	GTH: E: n: ANDEI	30 l ucle: DNES:	base ic a S: s:	pai: cid ingl		ur sepe	N 4 1 2	n Karaya		* + t t			ma, who
		(ii)) MO	LECU	LE T	YPE:	DNA	(ge	nomi	=)							
		(xi)) SE(QUEN	CE DI	ESCR:	IPTIC	ON:	SEQ 1	D NO	0:260):					

	ACAGGATCCA AATGCACAGC ACTGGTAAGC	30
	(262) INFORMATION FOR SEQ ID NO:261:	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 25 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:261:	
10	CTATAACTGG GTTACATGGT TTAAC	25
	(263) INFORMATION FOR SEQ ID NO:262:	
15	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:262:	
	TTTGAATTCA CATATTAATT AGAGACATGG	30
20	(264) INFORMATION FOR SEQ ID NO:263:	
25	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 2724 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:263:	
	ATGGACACCT CCCGGCTCGG TGTGCTCCTG TCCTTGCCTG TGCTGCTGCA GCTGGCGACC	60
	GGGGGCAGCT CTCCCAGGTC TGGTGTGTTG CTGAGGGGCT GCCCCACACA CTGTCATTGC	120
30	GAGCCCGACG GCAGGATGTT GCTCAGGGTG GACTGCTCCG ACCTGGGGCT CTCGGAGCTG	180
	CCTTCCAACC TCAGCGTCTT CACCTCCTAC CTAGACCTCA GTATGAACAA CATCAGTCAG	240
	CTGCTCCCGA ATCCCCTGCC CAGTCTCCGC TTCCTGGAGG AGTTACGTCT TGCGGGAAAC	300
	GCTCTGACAT ACATTCCCAA GGGAGCATTC ACTGGCCTTT ACAGTCTTAA AGTTCTTATG	360

CTGCAGAATA ATCAGCTAAG ACACGTACCC ACAGAAGCTC TGCAGAATTT GCGAAGCCTT CAATCCCTGC GTCTGGATGC TAACCACATC AGCTATGTGC CCCCAAGCTG TTTCAGTGGC 480 CTGCATTCCC TGAGGCACCT GTGGCTGGAT GACAATGCGT TAACAGAAAT CCCCGTCCAG 540 GCTTTTAGAA GTTTATCGGC ATTGCAAGCC ATGACCTTGG CCCTGAACAA AATACACCAC 600 ATACCAGACT ATGCCTTTGG AAACCTCTCC AGCTTGGTAG TTCTACATCT CCATAACAAT 660 AGAATCCACT CCCTGGGAAA GAAATGCTTT GATGGGCTCC ACAGCCTAGA GACTTTAGAT 720 TTAAATTACA ATAACCTTGA TGAATTCCCC ACTGCAATTA GGACACTCTC CAACCTTAAA 780 GAACTAGGAT TTCATAGCAA CAATATCAGG TCGATACCTG AGAAAGCATT TGTAGGCAAC 840 CCTTCTCTTA TTACAATACA TTTCTATGAC AATCCCATCC AATTTGTTGG GAGATCTGCT 900 TTTCAACATT TACCTGAACT AAGAACACTG ACTCTGAATG GTGCCTCACA AATAACTGAA 960 TTTCCTGATT TAACTGGAAC TGCAAACCTG GAGAGTCTGA CTTTAACTGG AGCACAGATC TCATCTCTTC CTCAAACCGT CTGCAATCAG TTACCTAATC TCCAAGTGCT AGATCTGTCT 1080 TACAACCTAT TAGAAGATTT ACCCAGTTTT TCAGTCTGCC AAAAGCTTCA GAAAATTGAC 1140 CTAAGACATA ATGAAATCTA CGAAATTAAA GTTGACACTT TCCAGCAGTT GCTTAGCCTC 1200 15 CGATCGCTGA ATTTGGCTTG GAACAAAATT GCTATTATTC ACCCCAATGC ATTTTCCACT 1260 TTGCCATCCC TAATAAAGCT GGACCTATCG TCCAACCTCC TGTCGTCTTT TCCTATAACT 1320 GGGTTACATG GTTTAACTCA CTTAAAATTA ACAGGAAATC ATGCCTTACA GAGCTTGATA 1380 TCATCTGAAA ACTTTCCAGA ACTCAAGGTT ATAGAAATGC CTTATGCTTA CCAGTGCTGT 1440 GCATTTGGAG TGTGTGAGAA TGCCTATAAG ATTTCTAATC AATGGAATAA AGGTGACAAC 1500 20 AGCAGTATGG ACGACCTTCA TAAGAAAGAT GCTGGAATGT TTCAGGCTCA AGATGAACGT 1560 GACCTTGAAG ATTTCCTGCT TGACTTTGAG GAAGACCTGA AAGCCCTTCA TTCAGTGCAG TGTTCACCTT CCCCAGGCCC CTTCAAACCC TGTGAACACC TGCTTGATGG CTGGCTGATC 1680 AGAATTGGAG TGTGGACCAT AGCAGTTCTG GCACTTACTT GTAATGCTTT GGTGACTTCA 1740 ACAGTTTCA GATCCCCTCT GTACATTCC CCCATTAAAC TGTTAATTGG GGTCATCGCA 1800 25 GCAGTGAACA TGCTCACGGG AGTCTCCAGT GCCGTGCTGG CTGGTGTGGA TGCGTTCACT 1860 TTTGGCAGCT TTGCACGACA TGGTGCCTGG TGGGAGAATG GGGTTGGTTG CCATGTCATT 1920 GGTTTTTTGT CCATTTTTGC TTCAGAATCA TCTGTTTTCC TGCTTACTCT GGCAGCCCTG 1980 GAGCGTGGGT TCTCTGTGAA ATATTCTGCA AAATTTGAAA CGAAAGCTCC ATTTTCTAGC 2040

BNSDOCID: <WO 002212941 1 5

	CTGA	LAAGT.	AA T	CATT	TTGC'	r ct	GTGC	CCTG	CTG	GCCT	TGA	CCAT	GGCC	GC A	GTTC	CCCT	G :	2100
	CTGG	GTGG	CA G	CAAG	TATG	G CG	CCTC	CCCT	CTC	TGCC	TGC	CTTT	GCCT	TT I	GGGG	AGCC	C :	2160
	AGCA	CCAT	GG G	CTAC.	ATGG'	r cg	CTCT	CATC	TTG	CTCA	ATT	CCCT	TTGC	TT C	CTCA	TGAT	G :	2220
	ACCA	TTGC	CT A	CACC.	AAGC'	г ст	ACTG	CAAT	TTG	GACA	AGG	GAGA	CCTG	GA G	AATA	TTTG	G :	2280
:	5 GACT	GCTC'	TA T	GGTA.	AAACI	A CA	TTGC	CCTG	TTG	CTCT	TCA	CCAA	CTGC	AT C	CTAA	ACTG	c a	2340
	CCTG	TGGC'	TT T	CTTG	TCCT	г ст	CCTC	TTTA	ATA	AACC'	TTA	CATT	TATC	AG T	CCTG.	AAGT.	A 2	2400
	ATTA	ÁGTT'	TA T	CCTT	CTGGT	r gg	TAGT	CCCA	CTT	CCTG	CAT	GTCT	CAAT	cc c	CTTC	TCTA	C 2	2460
	ATCT	TGTT	CA A	TCCT	CACT	TA.	AGGA	GGAT	CTG	GTGA	GCC	TGAG.	AAAG	CA A	ACCT	ACGT	C 2	520
	TGGA	CAAG	AT C	AAAA	CACC	AA	GCTT	GATG	TCA	ATTA	ACT	CTGA'	TGAT	FT C	GAAA	AACA	G 2	580
. 10) тсст	GTGA	CT C	AACT	CAAGO	CT	rggt	AACC	TTT	ACCA	GCT	CCAG	CATC	AC T	TATG	ACCT	G 2	640
	CCTC	CCAG	TT C	CGTG	CCATO	AC	CAGC'	TAT	CCA	GTGA	CTG .	AGAG	CTGC	CA T	CTTT	CCTC'	r 2	700
	GTGG	CATT	rg T	CCCA:	rgtci	CT	AA										2	724
	(265) INI	FORM	ATIOI	N FOR	SE	Q ID	NO:2	264:									
15	5	. (i)		_	CE CH					3								
			(B)	TYI	PE: a	min	ac:		-0.20.	•							•	
					POLOG			relev	ant									
		(ii	i) MO	DLECT	JLE 1	YPE	: pro	oteir	1									
. 20)	(xi	i) sı	EQUE	NCE I	ESCI	RIPT	ION:	SEQ	ID 1	10:2	64:						
		Met 1	Asp	Thr	Ser	Arg 5	Leu	Gly	Val	Leu	Leu 10	Ser	Leu	Pro	Val	Leu 15	Leu	
		Gln	Leu		Thr 20			Ser				Ser	Gly		Leu 30	Leu	Arg	
25	5	Gly	Cys	Pro 35	Thr	His	Çys	His	Cys 40	Glu	Pro	Asp	Gly	Arg 45	Met	Leu	Leu	
		Arg	Val 50	Asp	Cys	Ser	Asp	Leu 55	Gly	Leu	Ser	Glu	Leu 60	Pro	Ser	Asn	Leu	
30)	Ser 65	Val	Phe	Thr	Ser	Tyr 70	Leu	Asp	Leu	Ser	Met 75	Asn	Asn	Ile	Ser	Gln 80	
		Leu	Leu	Pro	Asn	Pro 85	Leu	Pro	Ser	Leu	Arg 90	Phe	Leu	Glu	Glu	Leu 95	Arg	

Leu Ala Gly Asn Ala Leu Thr Tyr Ile Pro Lys Gly Ala Phe Thr Gly

					100)				105	5				110)	
		Leu	Tyr	Ser 115	Leu	Lys	Val	Leu	Met 120	Leu	Glr.	Asn	Asn	Glr 125		Arg	, His
5	7	/al	Pro 130	Thr	Glu	Ala	Leu	Gln -135	Asn	. Leu	Arg	Ser	Leu 140	Gln		Leu	Arg
	1	eu 45	Asp	Ala	Asn	His	Ile 150	Ser	Tyr	Val	Pro	Pro 155		Cys	Phe	Ser	Gly 160
		eu	His	Ser	Leu	Arg 165	His	Leu	Trp	Leu	Asp 170			Ala	Leu	Thr 175	Glu
10	I	le	Pro	Val	Gln 180	Ala	Phe	Arg	Ser	Leu 185		Ala	Leu	Gln	Ala 190	Met	Thr
	I.	eu	Ala	Leu 195	Asn		Ile		His 200	Ile	Pro	Asp	Tyr	Ala 205	Phe	Gly	Asn
15	L	eu	Ser 210	Ser	Leu	Val	Val	Leu 215	His	Leu	His	Asn	Asn 220	Arg	Ile	His	Ser
	1. 2	eu 25	Gly	Lys	Lys	Суз	Phe 230	Asp	Gly	Leu	His	Ser 235	Leu	Glu	Thr	Leu	Asp 240
ı	L	eu	Asn	Tyr	Asn	Asn 245				Phe						Thr 255	Leu
20	s	er	Asn	Leu	Lys 260	Glu	Leu	Gly	Phe	His 265	Ser	Asn	Asn	Ile	Arg 270	Ser	Ile
	P	ro	Glu	Lys 275	Ala	Phe	Val		Asn 280			Leu		Thr 285		His	Phe
25	T ;	yr	Asp 290	Asn	Pro	Ile	Gln	Phe 295	Val	Gly			Ala 300	Phe		His	Leu
	P:	ro 05	Glu	Leu	Arg	Thr	Leu 310		Leu	Asn		Ala 315		Gln	Ile	Thr	Glu 320
	Pì	he	Pro	Asp	Leu	Thr 325	Gly	Thr	Ala	Asn	Leu 330	Glu	Ser	Leu	Thr	Leu 335	Thr
30	G	ly .	Ala	Gln	Ile 340	Ser	Ser	Leu	Pro	Gln 345	Thr	Val	Cys	Asn	Gln 350	Leu	Pro
	As	sn :	Leu	Gln 355	Val	Leu	qzA		Ser 360	Tyr	Asn	Leu	Leu	Glu 365	Asp	Leu	Pro
35	Se	er :	Phe 370	Ser	Val	Cys	Gln		Leu	Gln	Lys		Asp 380	Leu	Arg	His	Asn
	G]	lu : 35	Ile	Tyr	Glu		Lys 390	Val .	Asp	Thr		Gln 395	Gln	Leu ·	Leu		Leu 400

	•	Arg	Ser	Leu	Asn	Leu 405	Ala	Trp	Asn	Lys	Ile 410	Ala	Ile	Ile	His	Pro 415	Asn
		Ala	Phe	Ser	Thr 420	Leu	Pro	Ser	Leu	Ile 425	Lys	Leu	Asp	Leu	Ser 430	Ser	Asn
5		Leu	Leu	Ser 435	Ser	Phe	Pro	Ile	Thr 440	Gly	Leu	His	Gly	Leu 445	Thr	His	Leu
		Lys	Leu 450	Thr	Gly	Asn	His	Ala 455		Gln	Ser	Leu	Ile 460	Ser	Ser	Glu	Asn
10		Phe 465	Pro	Glu	Leu	Lys	Val 470	Ile	Glu	Met	Pro	Tyr 475	Ala	Tyr	Gln	Cys	Cys 480
		Ala	Phe	Gly	Val	Cys 485	Glu	Asn	Ala	Tyr	Lys 490	Ile	Ser	Asn	Gln	Trp 495	Asn
		Lys	Gly	Asp	Asn 500	Ser	Ser	Met	Asp	Asp 505	Leu	His	Lys	Lys	Asp 510	Ala	Gly
15		Met	Phe	Gln 515	Ala	Gln	Asp	Glu	Arg 520	Asp	Leu	Glu	Asp	Phe 525	Leu	Leu	Asp
		Phe	Glu 530	Glu	Asp	Leu	Lys	Ala 535	Leu	His	Ser	Val	Gln 540	Cys	Ser	Pro	Ser
20		Pro 545	Gly	Pro	Phe	Lys	Pro 550	Cys	Glu	His	Leu	Leu 555	Asp	Gly	Trp	Leu	Ile 560
		Arg	Ile	Gly	Val	Trp 565	Thr	Ile	Ala	Val	Leu 570	Ala	Leu	Thr	Cys	Asn 575	Ala
		Leu	Val	Thr	Ser 580	Thr	Val	Phe	Arg	Ser 585	Pro	Leu	Tyr	Ile	Ser 590	Pro	Ile
25		Lys	Leu	Leu 595	Ile	Gly	Val	Ile	Ala 600	Ala	Val	Asn	Met	Leu 605	Thr	Gly	Val
		Ser	Ser 610	Ala	Val	Leu	Ala	Gly 615	Val	Asp	Ala	Phe	Thr 620	Phe	Gly	Ser	Phe
30		Ala 625	Arg	His	Gly	Ala	Trp 630	Trp	Glu	Asn	Gly	Val 635	Gly	Cys	His	Val	Ile 640
		Gly	Phe	Leu	Ser	Ile 645	Phe	Ala	Ser	Glu	Ser 650	Ser	Val	Phe	Leu	Leu 655	Thr
		Leu	Ala	Ala	Leu 660	Glu	Arg	Gly	Phe	Ser 665	Val	Lys	Tyr	Ser	Ala 670	Lys	Phe
35		Glu	Thr	Lys 675	Ala	Pro	Phe	Ser	Ser 680	Leu	Lys	Val	Ile	Ile 685	Leu	Leu	Cys
		Ala	Leu	Leu	Ala	Leu	Thr	Met	Ala	Ala	Val	Pro	Leu	Leu	Gly	Gly	Ser

Committee to the state of the

			690				:	695					700				
		Lys 705	Tyr	Gly	Ala	Ser	Pro 710	Leu	Суз	Leu	Pro	Leu 715	Pro	Phe	Gly	Glu	Pro 720
5		Ser	Thr	Met	Gly	Tyr 725	Met	Val	Ala	Leu	Ile 730	Leu	Leu	Asn	Ser	Leu 735	Cys
		Phe	Leu	Met	Met 740	Thr	Ile	Ala	Tyr	Thr 745	Lys	Leu	Tyr	Cys	Asn 750	Leu	Asp
		Lys	Gly	Asp 755	Leu	Glu	Asn	Ile	Trp 760	Asp	Cys	Ser	Met	Val 765	Lys	His	Ile
10		Ala	Leu 770	Leu	Leu	Phe	Thr	Asn 775	Cys	Ile	Leu	Asn	Cys 780	Pro	Val	Ala	Phe
		Leu 785	Ser	Phe	Ser	Ser	Leu 790	Ile	Asn	Leu	Thr	Phe 795	Ile	Ser	Pro	Glu	Val 800
15		Ile	Lys	Phe	Ile	Leu 805	Leu	Val	Val	Val	Pro 810	Leu	Pro	Ala	Cys	Leu 815	Asn
		Pro	Leu	Leu	Tyr 820	Ile	Leu	Phe	Asn	Pro 825	His	Phe	Lys		Asp 830	Leu	Val
		Ser	Leu	Arg 835	Lys	Gln	Thr	Tyr	Val 840	Trp	Thr	Arg	Ser	Lys 845	His	Pro	Ser
20		Leu	Met 850	Ser	Ile	Asn	Ser	Asp 855	Asp	Val	Glu	Lys	Gln 860	Ser	Cys	Asp	Ser
		Thr 865	Gln	Ala	Leu	Val	Thr 870	Phe	Thr	Ser	Ser	Ser 875	Ile	Thr	Tyr		Leu 880
25		Pro	Pro	Ser	Ser	Val 885	Pro	Ser	Pro		Tyr 890	Pro	Val	Thr		Ser 895	Cys
	1	His	Leu		Ser 900	Val	Ala	Phe	Val	Pro 905	Cys	Leu					
	(266)	INF	ORMA	TION	FOR	SEQ	ID	NO:2	65:					,			
30		(i)	(A) (B) (C)	TYP.	GTH: E: n ANDE	ARAC 30 ucle DNES Y: 1	base ic a S: s	pai cid ingl	rs							٠	
•		(ii				YPE:			nomi	c)							
35		(xi) SE	QUEN	CE D	ESCR	IPTI	ON:	SEQ :	ID N	0:26	5 :					
	CGGAA	SCTG	C GG	GCCA	aatg	GGT	GGCC	GGC									30

	(267) INFORMATION FOR SEQ ID NO:266:	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 27 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic) (xi) SEQUENCE DESCRIPTION: SEQ ID NO:266:	
	CAGAGGAGGG TGAAGGGGCT GTTGGCG	27
10	(268) INFORMATION FOR SEQ ID NO:267:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:267:	•
	GGCGGCGCG AGCCAAGGGG CTGGCTGTGG	30
	(269) INFORMATION FOR SEQ ID NO:268:	
20	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 32 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
25	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:268:	
	GGGACTGCTC TATGAAAAAA CACATTGCCC TG	32
	(270) INFORMATION FOR SEQ ID NO:269:	
30	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1071 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:269:	
	ATGAATGGGG TCTCGGAGGG GACCAGAGGC TGCAGTGACA GGCAACCTGG GGTCCTGACA	60

	CGTGATCGCT CTTGTTCCAG GAAGATGAAC TCTTCCGGAT GCCTGTCTGA GGAGGTGGGG	120
	TCCCTCCGCC CACTGACTGT GGTTATCCTG TCTGCGTCCA TTGTCGTCGG AGTGCTGGGC	180
	AATGGGCTGG TGCTGTGGAT GACTGTCTTC CGTATGGCAC GCACGGTCTC CACCGTCTGC	240
	TTCTTCCACC TGGCCCTTGC CGATTTCATG CTCTCACTGT CTCTGCCCAT TGCCATGTAC	300
5	TATATTGTCT CCAGGCAGTG GCTCCTCGGA GAGTGGGCCT GCAAACTCTA CATCACCTTT	360
	GTGTTCCTCA GCTACTTTGC CAGTAACTGC CTCCTTGTCT TCATCTCTGT GGACCGTTGC	420
•	ATCTCTGTCC TCTACCCCGT CTGGGCCCTG AACCACCGCA CTGTGCAGCG GGCGAGCTGG	480
	CTGGCCTTTG GGGTGTGGCT CCTGGCCGCC GCCTTGTGCT CTGCGCACCT GAAATTCCGG	540
	ACAACCAGAA AATGGAATGG CTGTACGCAC TGCTACTTGG CGTTCAACTC TGACAATGAG	600
10	ACTGCCCAGA TTTGGATTGA AGGGGTCGTG GAGGGACACA TTATAGGGAC CATTGGCCAC	660
	TTCCTGCTGG GCTTCCTGGG GCCCTTAGCA ATCATAGGCA CCTGCGCCCA CCTCATCCGG	720
	GCCAAGCTCT TGCGGGAGGG CTGGGTCCAT GCCAACCGGC CCAAGAGGCT GCTGCTGGTG	780
	CTGGTGAGCG CTTTCTTTAT CTTCTGGTCC CCGTTTAACG TGGTGCTGTT GGTCCATCTG	840
,	TGGCGACGGG TGATGCTCAA GGAAATCTAC CACCCCGGA TGCTGCTCAT CCTCCAGGCT	900
15	AGCTTTGCCT TGGGCTGTGT CAACAGCAGC CTCAACCCCT TCCTCTACGT CTTCGTTGGC	960
	AGAGATTTCC AAGAAAAGTT TTTCCAGTCT TTGACTTCTG CCCTGGCGAG GGCGTTTGGA 1	020
,	GAGGAGGAGT TTCTGTCATC CTGTCCCCGT GGCAACGCCC CCCGGGAATG A	071
	(271) INFORMATION FOR SEQ ID NO:270:	
20	(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 356 amino acids(B) TYPE: amino acid	
	(C) STRANDEDNESS: (D) TOPOLOGY: not relevant	
	(ii) MOLECULE TYPE: protein	
25	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:270:	
	Met Asn Gly Val Ser Glu Gly Thr Arg Gly Cys Ser Asn Arg Gla Bro	

Gly Val Leu Thr Arg Asp Arg Ser Cys Ser Arg Lys Met Asn Ser Ser

Gly Cys Leu Ser Glu Glu Val Gly Ser Leu Arg Pro Leu Thr Val Val

35 40 45

יו וישטניטטי שואר שואר שואנים

-	Ile	Leu 50	Ser	Ala	Ser	Ile	Val 55	Val	Gly	Val	Leu	Gly 60	Asn	Gly	Leu	Val
	Leu 65	Trp	Met	Thr	Val	Phe 70	Arg	Met	Ala	Arg	Thr 75	Val	Ser	Thr	Val	Суs 80
5	Phe	Phe	His	Leu	Ala 85	Leu	Ala	Asp	Phe	Met 90	Leu	Ser	Leu	Ser	Leu 95	Pro
	Ile	Ala	Met	Tyr 100	Tyr	Ile	Val	Ser	Arg 105	Gln	Trp	Leu	Leu	Gly 110	Glu	Trp
10	Ala	Cys	Lys 115	Leu	Tyr	Ile	Thr	Phe 120	Val	Phe	Leu	Ser	Tyr 125	Phe	Ala	Ser
	Asn	Cys 130	Leu	Leu	Val	Phe	Ile 135	Ser	Val	Asp	Arg	Cys 140	Ile	Ser	Val	Leu
	Tyr 145	Pro	Val	Trp	Ala	Leu 150	Asn	His	Arg	Thr	Val 155	Gln	Arg	Ala	Ser	Trp 160
15	Leu	Ala	Phe	Gly	Val 165	Trp	Leu	Leu	Ala	Ala 170	Ala	Leu	Cys	Ser	Ala 175	His
	Leu	Lys	Phe	Arg 180	Thr	Thr	Arg	Lys	Trp 185	Asn	Gly	Cys	Thr	His 190	Cys	Tyr
20	Leu	Ala	Phe 195	Asn	Ser	Asp	Asn	Glu 200	Thr	Ala	Gln	Ile	Trp 205	Ile	Glu	Gly
	Val	Val 210	Glu	Gly	His	Ile	Ile 215	Gly	Thr	Ile	Gly	His 220	Phe	Leu	Leu	Gly
	Phe 225	Leu	Gly	Pro	Leu	Ala 230	Ile	Ile	Gly	Thr	Cys 235	Ala	His	Leu	Ile	Arg 240
25	Ala	Lys	Leu	Leu	Arg 245	Glu	Gly	Trp	Val	His 250	Ala	Asn	Arg	Pro	Lys 255	Arg
	Leu	Leu	Leu	Val 260	Leu	Val	Ser	Ala	Phe 265	Phe	Ile	Phe	Trp	Ser 270	Pro	Phe
30	Asn	Val	Val 275	Leu	Leu	Val	His	Leu 280	Trp	Arg	Arg	Val	Met 285	Leu	Lys	Glu
	Ile	Tyr 290	His	Pro	Arg	Met	Leu 295	Leu	Ile	Leu	Gln	Ala 300	Ser	Phe	Ala	Leu
	Gly 305	Cys	Val	Asn	Ser	Ser 310	Leu	Asn	Pro	Phe	Leu 315	Tyr	Val	Phe	Val	Gly 320
35	Arg	Asp	Phe	Gln	Glu 325	Lys	Phe	Phe	Gln	Ser 330	Leu	Thr	Ser	Ala	Leu 335	Ala
	Arg	Ala	Phe	Gly	Glu	Glu	Glu	Phe	Leu	Ser	Ser	Cys	Pro	Arg	Gly	Asn

60

180

240

300

420

480

540

600

660

780

840

900

903

340

(272) INFORMATION FOR SEQ ID NO:271:

(B) TYPE: nucleic acid

State of the second second

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 903 base pairs

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:271:

ATGGACCTGC CCCCGCAGCT CTCCTTCGGC CTCTATGTGG CCGCCTTTGC GCTGGGCTTC

CCGCTCAACG TCCTGGCCAT CCGAGGCGCG ACGGCCCACG CCCGGCTCCG TCTCACCCCT

CTGAAGGCGG TGGAGGCGCT AGCCTCCGGG GCCTGGCCTC TGCCGGCCTC GCTGTGCCCC

GTCTTCGCGG TGGCCCACTT CTTCCCACTC TATGCCGGCG GGGGCTTCCT GGCCGCCCTG

15. AGCCTGGTCT ACGCCCTGAA CCTGGGCTGC TCCGACCTGC TGCTGACAGT CTCTCTGCCC

Ala Pro Arg Glu 355

5

10

M2212041 1 -

DINY GIOCUSIAE

.:-	AGTGCAGGCC	GCTACCTGGG	AGCAGCCTTC	CCCTTGGGCT	ACCAAGCCTT	CCGGAGGCCG
	TGCTATTCCT	GGGGGGTGTG	CGCGGCCATC	TGGGCCCTCG	TCCTGTGTCA	CCTGGGTCTG
20	GTCTTTGGGT	TGGAGGCTCC	AGGAGGCTGG	CTGGACCACA	GCAACACCTC	CCTGGGCATC
	AACACACCGG	TCAACGGCTC	TCCGGTCTGC	CTGGAGGCCT	GGGACCCGGC	CTCTGCCGGC
. ;	CCGGCCCGCT	TCAGCCTCTC	TCTCCTGCTC	TTTTTTCTGC	CCTTGGCCAT	CACAGCCTTC
	TGCTACGTGG	GCTGCCTCCG	GGCACTGGCC	CGCTCCGGCC	TGACGCACAG	GCGGAAGCTG
	CGGGCCAAAT	GGGTGGCCGG	CGGGGCCCTC	CTCACGCTGC	TGCTCTGCGT	AGGACCCTAC
25	AACGCCTCCA	ACGTGGCCAG	CTTCCTGTAC	CCCAATCTAG	GAGGCTCCTG	GCGGAAGCTG
	GGGCTCATCA	CGGGTGCCTG	GAGTGTGGTG	CTTAATCCGC	TGGTGACCGG	TTACTTGGGA
	AGGGGTCCTG	GCCTGAAGAC	AGTGTGTGCG	GCAAGAACGC	AAGGGGGCAA	GTCCCAGAAG
	TAA					
	(273) INFOR	MATION FOR	SEQ ID NO:2	272:		: :
30		EQUENCE CHA			•	
		B) TYPE: an C) STRANDEI		e ver e e	e e e e e e e e e e e e e e e e e e e	

(D)	TOPOLOGY:	not	relevant
-----	-----------	-----	----------

- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:272:

	-	•	-					_								
5	Met 1	Asp	Leu	Pro	Pro 5	Gln	Leu	Ser	Phe	Gly 10	Leu	Tyr	Val	Ala	Ala 15	Phe
	Ala	Leu	Gly	Phe 20	Pro	Leu	Asn	Val	Leu 25	Ala	Ile	Arg	Gly	Ala 30	Thr	Ala
	His	Ala	Arg 35	Leu	Arg	Leu	Thr	Pro 40	Ser	Leu	Val	Tyr	Ala 45	Leu	Asn	Leu
10	Gly	Сув 50	Ser	Asp	Leu	Leu	Leu 55	Thr	Val	Ser	Leu	Pro 60	Leu	Lys	Ala	Val
	Glu 65	Ala	Leu	Ala	Ser	Gly 70	Ala	Trp	Pro	Leu	Pro 75	Ala	Ser	Leu	Cys	Pro 80
15	Val	Phe	Ala	Val	Ala 85	His	Phe	Phe	Pro	Leu 90	Tyr	Ala	Gly	Gly	Gly 95	Phe ,
	Leu	Ala	Ala	Leu 100	Ser	Ala	Gly	Arg	Tyr 105	Leu	Gly	Ala	Ala	Phe 110	Pro	Leu
	Gly	Tyr	Gln 115	Ala	Phe	Arg	Arg	Pro 120	Cys	Tyr	Ser	Trp	Gly 125	Val	Cys	Ala;
20	Ala	Ile 130	Trp	Ala	Leu	Val	Leu 135	Cys	His	Leu	Gly	Leu 140	Val	Phe	Gly	Leu
•	Glu 145	Ala	Pro	Gly	Gly	Trp 150	Leu	Asp	His	Ser	Asn 155	Thr	Ser	Leu	Gly	Ile 160
25	Asn	Thr	Pro	Val	Asn 165	Gly	Ser	Pro	Val	Cys 170	Leu	Glu	Ala	Trp	Asp 175	Pro
·	Ala	Ser	Ala	Gly 180	Pro	Ala	Arg	Phe	Ser 185	Leu	Ser	Leu	Leu	Leu 190	Phe	Phe
	Leu	Pro	Leu 195	Ala	Ile	Thr	Ala	Phe 200	Cys	Tyr	Val	Gly	Cys 205	Leu	Arg	Ala
30	Leu	Ala 210	Arg	Ser	Gly	Leu	Thr 215	His	Arg	Arg	Lys	Leu 220	Arg	Ala	Lys	Trp
	Val 225	Ala	Gly	Gly	Ala	Leu 230		Thr	Leu	Leu	Leu 235	Cys	Val	Gly	Pro	Tyr 240
35	Asn	Ala	Ser	Asn	Val 245	Ala	Ser	Phe	Leu	Tyr 250	Pro	Asn	Leu	Gly	Gly 255	Ser
	Trp	Arg	Lys	Leu	Gly	Leu	Ile	Thr	Gly	Ala	Trp	Ser	Val	Val	Leu	Asn

		260	÷ #	265		. 270	
	Pro L	eu Val Thr 275	Gly Tyr Leu	Gly Arg Gl 280		Leu Lys Thr	Val
5		la Ala Arg 90	Thr Gln Gly 295				
	(274). INFO	RMATION FOR	SEQ ID NO:	273:	•		
10	(i)	(B) TYPE: n	1041 base pucleic acid DNESS: sing	pairs	er e e		
		MOLECULE T			273:		
	ATGGATACAG	GCCCGACCA	GTCCTACTTC	TCCGGCAATC	ACTGGTTCGT	CTTCTCGGTG	60
15	TACCTTCTCA	CTTTCCTGGT	GGGCTCCCC	CTCAACCTGC	TGGCCCTGGT	GGTCTTCGTG	120
	GGCAAGCTGC	AGCGCCGCCC	GGTGGCCGTG	GACGTGCTCC	TGCTCAACCT	GACCGCCTCG	180
	GACCTGCTCC	TGCTGCTGTT	CCTGCCTTTC	CGCATGGTGG	AGGCAGCCAA	TGGCATGCAC	240
9 9	TGGCCCCTGC	CCTTCATCCT	CTGCCCACTC	TCTGGATTCA	TCTTCTTCAC	CACCATCTAT	300
	CTCACCGCCC	TCTTCCTGGC	AGCTGTGAGC	ATTGAACGCT	TCCTGAGTGT	GGCCCACCCA	360
20	CTGTGGTACA	AGACCCGGCC	GAGGCTGGGG	CAGGCAGGTC	TGGTGAGTGT	GGCCTGCTGG	420
	CTGTTGGCCT	CTGCTCACTG	CAGCGTGGTC	TACGTCATAG	AATTCTCAGG	GGACATCTCC	480
	CACAGCCAGG	GCACCAATGG	GACCTGCTAC	CTGGAGTTCC	GGAAGGACCA	GCTAGCCATC	540
	CTCCTGCCCG	TGCGGCTGGA	GATGGCTGTG	GTCCTCTTTG	TGGTCCCGCT	GATCATCACC	600
	AGCTACTGCT	ACAGCCGCCT	GGTGTGGATC	CTCGGCAGAG	GGGGCAGCCA	CCGCCGGCAG	660
25	AGGAGGGTGA	AGGGGCTGTT	GGCGGCCACG	CTGCTCAACT	TCCTTGTCTG	CTTTGGGCCC	720
•	TACAACGTGT	CCCATGTCGT	GGGCTATATC	TGCGGTGAAA	GCCCGGCATG	GAGGATCTAC	780
	GTGACGCTTC	TCAGCACCCT	GAACTCCTGT	GTCGACCCCT	TTGTCTACTA	CTTCTCCTCC	840
	TCCGGGTTCC	AAGCCGACTT	TCATGAGCTG	CTGAGGAGGT	TGTGTGGGCT	CTGGGGCCAG	900
•	TGGCAGCAGG	AGAGCAGCAT	GGAGCTGAAG	GAGCAGAAGG	GAGGGGAGGA	GCAGAGAGCG	960
30	GACCGACCAG	CTGAAAGAAA	GACCAGTGAA	CACTCACAGG	GCTGTGGAAC	TGGTGGCCAG	1020
	GTGGCCTGTG	CTGAAAGCTA	G				1041

South the the there's

227

(275) INFORMATION FOR SEQ ID NO:274:

5		(i	(A (B (C) LE) TY) ST	NGTH PE: RAND	HARA : 34 amin EDNE	6 am o ac SS:	ino i	acid	s	••						
		(i:	i) M	OLEC	ULE '	TYPE	: pr	otei	n ··			*					
		(x:	i) s	EQUE	NCE I	DESC	RIPT	ION:	SEQ	ID :	NO:2	74:	-				
10		Met 1	Asp	Thr	Gly	Pro 5	Asp	Gln	Ser	Tyr	Phe 10	Ser	Gly	Asn	His	Trp 15	Phe
		Val	Phe	Ser	Val 20	Tyr	Leu	Leu	Thr	Phe 25	Leu	Val	Gly	Leu	Pro 30	Leu	Asn
		Leu	Leu	Ala 35	Leu	Val	Val	Phe	Val 40	Gly	Lys	Leu	Gln	Arg 45	Arg	Pro	Val
15		Ala	Val 50	Asp	Val	Leu	Leu	Leu 55	Asn	Leu	Thr	Ala	Ser 60	Asp	Leu	Leu	Leu
	•	Leu 65	Leu	Phe	Leu	Pro	Phe 70	Arg	Met	Val	Glu	Ala 75	Ala	Asn	Gly	Met	His 80
20		Trp	.Pro	Leu	Pro	Phe 85	Ile	Leu	Cys	Pro	Leu 90	Ser	Gly	Phe	Ile	Phe 95	Phe
	•	Thr	Thr	Ile	Tyr 100	Leu	Thr	Ala	Leu	Phe 105	Leu	Ala	Ala	Val	Ser 110	Ile	Glu
		Arg	Phe	Leu 115	Ser	Val	Ala	His	Pro 120	Leu	Trp	Tyr	Lys	Thr 125	Arg	Pro	Arg
25	•	Leu	Gly 130	Gln	Ala	Gly	Leu	Val 135	Ser	Val	Ala	Cys	Trp 140	Leu	Leu	Ala	Ser
		Ala 145	His	Cys	Ser	Val	Val 150	Tyr	Val	Ile	Glu	Phe 155	Ser	Gly	Asp	Ile	Ser 160
30	•	His	Ser	Gln		Thr 165	Asn	Gly	Thr	Cys	Туr 170	Leu	Glu	Phe	Arg	Lys 175	Asp
		Gln	Leu	Ala	Ile 180	Leu	Leu	Pro	Val	Arg 185	Leu	Glu	Met	Ala	Val 190	Val	Leu
		Phe	Val	Val 195	Pro	Leu	Ile	Ile	Thr 200	Ser	Tyr	Cys	Tyr	Ser 205	Arg	Leu	Val
35		Trp	Ile 210	Leu	Gly	Arg	Gly	Gly 215	Ser	His	Arg	Arg	Gln 220	Arg	Arg	Val	Lys
		Gly	Leu	Leu	Ala	Ala	Thr	Leu	Leu	Asn	Phe	Leu	Val	Cys	Phe	Gly	Pro

	225					230					235					240	
	Tyr	Asn	Val	Ser	His 245	Val	Val	Gly	Tyr	Ile 250	Cys	Gly	Glu	Ser	Pro 255	Ala	
5	Trp	Arg	Ile	Tyr 260		Thr	Leu	Leu	Ser 265		Leu			Cys 270		Asp	
	Pro	Phe	Val 275	Tyr	Tyr	Phe	Ser	Ser 280	Ser	Gly	Phe	Gln	Ala 285	Asp	Phe	His	
	Glu	Leu 290	Leu	Arg	Arg	Leu	Cys 295	Gly	Leu	Trp	Gly	Gln 300	Trp	Gln	Gln	Glu	
10	Ser 305	Ser	Met	Glu	Leu	Lys 310	Glu	Gln	Lys	Gly	Gly 315	Glu	Glu	Gln	Arg	Ala 320	
	Asp	Arg	Pro	Ala	Glu 325	Arg	Lys	Thr	Ser	Glu 330	His	Ser	Gln	Gly	Cys 335	Gly	
15	Thr	Gly	Gly	Gln 340	Val	Ala	Cys	Ala	Glu 345	Ser		• .	**				
	(276) IN	FORMA	TION	1 FOF	SEÇ	Q ID	NO:2	75 :									
	(i)		LEN	IGTH :	993	bas	se pa										
20		(C)	STR	PE: r RANDE POLOG	DNES	SS: s	singl	.e									
	(i :	i) MC	LECU	JLE I	YPE:	DNA	ı (ge	nomi	.c)								
	(xi	i) SE	QUEN	ICE I	ESCF	RIPTI	ON:	SEQ	ID N	10:27	75:						
	ATGCTGCCC	GG AC	TGGA	AGAG	CTC	CTT	ATC	CTCA	TGGC	TT A	CATO	ATCA	T CI	TCCI	CACI	•	60
25	GGCCTCCCT	rg cc	'AACC	TCCI	GGC	CCT	CGG	GCCI	TTGI	GG G	GCGG	ATC	G, CC	CAGCO	CCAG		L20
	CCTGCACC	rg tg	CACA	ATCCI	CCI	GCT	BAGC	CTGA	CGCI	GG C	CGAC	CTCC	T CC	TGCI	GCTG	. 1	L80
	CTGCTGCC	CT TO	:AAGA	TCAT	CGA	AGGCI	rgcg	TCGA	ACTI	CC G	CTGC	TACC	T GC	CCAP	AGGTO	: 2	240
	GTCTGCGC	CC TC	ACGA	GTTI	TGG	CTT	TAC	AGCA	GCAT	CT A	CTGC	AGCA	C GI	GGCI	CCTG	; 3	300
	GCGGGCAT	CA GC	ATC	AGCG	CTA	ACCTO	GGA	GTGG	CTTI	ec c	CGTG	CAGI	'A CA	AGCI	CTCC	. 3	360
30	CGCCGGCCT	rc TG	TATG	GAGI	GAT	TGC	AGCT	CTGG	TGGC	CT G	GGTI	ATGI	C CI	TTG	STCAC	. 4	120
	TGCACCATO	CG TG	ATCA	TCGI	TCA	\ATA(TTG	AACA	CGAC	TG A	GCAG	GTCA	G AA	GTG	CAAT	. 4	80
	GAAATTAC	CT GC	TACG	AGAA	CTI	CAC	GAT	AACC	AGTI	GG A	CGTG	GTGC	T GC	CCGI	rgcge	; 5	540
	ריזינוב א ביריזינ	פיי פיי	יריתינים	יייברייו	ئىشە د	יריייי	יאיזירי	cca	maaa	יארי ת	 '''	יא פורות		10 0m	amaa		- 0 0

• •	CGTTTTGTC	FT G	GATC	ATGC'	r ct	CCCA	.GCCC	CTI	GTGG	GGG	CCCA	GAGG	CG G	CGCC	GAGC	C.	660
	AAGGGGCTG	G C	TGTG	GTGA	C GC	TGCT	CAAT	TTC	CTGG	TGT	GCTI	CGGA	CC 1	TACA	ACGI	Ğ	720
	TCCCACCTC	G T	GGGG	TATC	A CC	AGAG	AAAA	AGC	CCCT	GGT	GGCG	GTCA	AT A	GCCG	TGGT	Ğ	780
	TTCAGTTCA	C T	CAAC	GCCA	G TC	TGGA	cccc	CTG	CTCT	TCT	ATTT	CTCT	TC I	TCAG	TGGT	'G	840
5	CGCAGGGCA	T	rggg:	AGAG	GC	TGCA	GGTG	CTG	CGGA	ATC	AGGG	CTCC	TC C	CTGT	TGGG	A	900
	CGCAGAGGC	A A	AGAC	ACAG	C AG	AGGG	GACA	AAT	GAGG	ACA	GGGG	TGTG	GG I	CAAG	GAGA	A	960
	GGGATGCCA	A G	TTCG	GACT	CA	CTAC	AGAG	TAG									993
	(277) INF	ORM	ATIO	N FOI	SE	Q ID	NO:	276:			•						
10		(A)	LEI	CE CH NGTH: PE: & RANDI	33 min	0 am	ino a		s								
	(ii			DLE 1					-								
15	(xi) SI	EQUE	NCE I	ESC	RIPT:	ION:	SEQ	ID I	NO:2	76 :						
	Met 1	Leu	Pro	Asp	Trp 5	Lys	Ser	Ser	Leu	Ile 10	Leu	Met	Ala	Tyr	Ile 15	Ile	
	Ile	Phe	Leu	Thr 20	Gly	Leu	Pro	Ala	Asn 25	Leu	Leu	Ala	Leu	Arg 30	Ala	Phe	:
20	Val	Gly	Arg 35	Ile	Arg	Gln	Pro	Gln 40	Pro	Ala	Pro	Val	His 45	Ile	Leu	Leu	*.: *
	Leu	Ser 50	Leu	Thr	Leu	Ala	Asp 55	Leu	Leu	Leu	Leu	Leu 60	Leu	Leu	Pro	Phe	
25	Lys 65	Ile	Ile	Glu	Ala	Ala 70	Ser	Asn	Phe	Arg	Trp 75	Tyr	Leu	Pro	Lys	Val 80	
	Val (Cys	Ala	Leu	Thr 85	Ser	Phe	Gly	Phe	Tyr 90	Ser	Ser	Ile	Tyr	Cys 95	Ser	
	Thr '	Trp	Leu	Leu 100	Ala	Gly	Ile	Ser	Ile 105	Glu	Arg	Tyr	Leu	Gly 110	Val	Ala	
30	Phe :	Pro	Val 115	Gln	Tyr	Lys	Leu	Ser 120	Arg	Arg	Pro	Leu	Tyr 125	Gly	Val	Ile	
	Ala 2	Ala 130	Leu	Val	Ala	Trp	Val 135	Met	Ser	Phe	Gly	His 140	Cys	Thr	Ile	Val	
35	Ile : 145	Ile	Val	Gln	Tyr	Leu 150	Asn	Thr	Thr	Glu	Gln 155	Val	Arg	Ser	Gly	Asn 160	

		Glu	Ile	Thr	Cys	Tyr 165	Glu	Asn	Phe	Thr	Asp 170	Asn	Gln	Leu	Asp	Val 175	Val
	٠	Leu	Pro	Val	Arg 180	Leu	Glu	Leu	Cys	Leu 185	Val	Leu	Phe	Phe	Ile 190	Pro	Met
. 5		Ala	Val	Thr 195	Ile	Phe	Cys	Tyr	Trp 200	Arg	Phe	Val	Trp	Ile 205	Met	Leu	Ser
		Gln	Pro 210	Leu	Val	Gly	Ala	Gln 215	Arg	Arg	Arg	Arg	Ala 220	Lys	Gly	Leu	Ala
10		Val 225	Val	Thr	Leu	Leu	Asn 230	Phe	Leu	Val	Cys	Phe 235	Gly	Pro	Tyr	Asn	Val 240
		Ser	His	Leu	Val	Gly 245	Tyr	His	Gln	Arg	Lys 250	Ser	Pro	Trp	Trp	Arg 255	Ser
	•	Ile.	Ala	Val	Val 260	Phe	Ser	Ser	Leu	Asn 265	Ala	Ser	Leu	Asp	Pro 270	Leu	Leu
15		Phe	Tyr	Phe 275	Ser	Ser	Ser	Val	Val 280	Arg	Arg	Ala	Phe	Gly 285	Arg	Gly	Leu
		Gln	Val 290	Leu	Arg	Asn	Gln	Gly 295	Ser	Ser	Leu		Gly 300	Arg	Arg	Gly	Lys
20		Asp 305	Thr	Ala	Glu	Gly	Thr 310	Asn	Glu	Asp		Gly 315	Val	Gly	Gln		Glu 320
		Gly	Met	Pro	Ser	Ser 325	Asp	Phe	Thr		Glu 330			٠			
	(278)	INF	'ORMA	TION	FOR	SEQ	ID	NO:2	77:								
25			(A) (B) (C) (D)	LEN TYP STR TOP	E CH GTH: E: n ANDE	272 ucle DNES Y: 1	4 ba ic a S: s inea	se p cid ingl r	airs e							•	
		(ii) MO	LECU	LE T	YPE:	DNA	(ge	nomi	c)							
30		(xi) SE	QUEN	CE D	ESCR	IPTI	ON:	SEQ	ID N	0:27	7:					
	ATGGA	CACC'	T CC	CGGC	TCGG	TGT	GCTC	CTG	TCCT	TGCC'	TG T	GCTG	CTGC	A GC	rggc	GACC	60
	GGGGG	CAGC'	T CT	CCCA	GGTC	TGG'	TGTG:	TTG (CTGA	GGGG	CT G	CCCI	ACAC	A CT	GTCA:	TGC	120
	GAGCC	CGAC	G GC	AGGA'	TGTT	GCT	CAGG	GTG (GACT	GCTC	CG A	CCTGC	GGC:	r ct	CGGA	CTG	180
	CCTTC																240
35	CTGCT	CCCG	TA A	CCCC'	rgcc	CAG	rctco	CGC 1	rtcc:	rggac	G AC	STTAC	GTCI	TG	:GGG	LAAC	300

PCT/US99/23938

231

WO 00/22129

• .	GCTCTGACAT	ACATTCCCAA	GGGAGCATTC	ACTGGCCTTT	ACAGTCTTAA	AGTTCTTATG	360
	CTGCAGAATA	ATCAGCTAAG	ACACGTACCC	ACAGAAGCTC	TGCAGAATTT	GCGAAGCCTT	420
	CAATCCCTGC	GTCTGGATGC	TAACCACATC	AGCTATGTGC	CCCCAAGCTG	TTTCAGTGGC	400
	CTGCATTCCC	TGAGGCACCT	GTGGCTGGAT	GACAATGCGT	TAACAGAAAT	CCCCGTCCAG	540
5	GCTTTTAGAA	GTTTATCGGC	ATTGCAAGCC	ATGACCTTGG	CCCTGAACAA	AATACACCAC	600
	ATACCAGACT	ATGCCTTTGG	AAACCTCTCC	AGCTTGGTAG	TTCTACATCT	CCATAACAAT	660
	AGAATCCACT	CCCTGGGAAA	GAAATGCTTT	GATGGGCTCC	ACAGCCTAGA	GACTTTAGAT	720
	TTAAATTACA	ATAACCTTGA	TGAATTCCCC	ACTGCAATTA	GGACACTCTC	CAACCTTAAA	780
	GAACTAGGAT	TTCATAGCAA	CAATATCAGG	TCGATACCTG	AGAAAGCATT	TGTAGGCAAC	840
10	CCTTCTCTTA	TTACAATACA	TTTCTATGAC	AATCCCATCC	AATTTGTTGG	GAGATCTGCT	900
	TTTCAACATT	TACCTGAACT	AAGAACACTG	ACTCTGAATG	GTGCCTCACA	AATAACTGAA	960
	TTTCCTGATT	TAACTGGAAC	TGCAAACCTG	GAGAGTCTGA	CTTTAACTGG	AGCACAGATC	1020
	TCATCTCTTC	CTCAAACCGT	CTGCAATCAG	TTACCTAATC	TCCAAGTGCT	AGATCTGTCT .	1080
	TACAACCTAT	TAGAAGATTT	ACCCAGTTTT	TCAGTCTGCC	AAAAGCTTCA	GAAAATTGAC	1140
15	CTAAGACATA	ATGAAATCTA	CGAAATTAAA	GTTGACACTT	TCCAGCAGTT	GCTTAGCCTC	1200
	CGATCGCTGA	ATTTGGCTTG	GAACAAAATT	GCTATTATTC	ACCCCAATGC	ATTTTCCACT	1260
	TTGCCATCCC	TAATAAAGCT	GGACCTATCG	TCCAACCTCC	TGTCGTCTTT	TCCTATAACT	1320
	GGGTTACATG	GTTTAACTCA	CTTAAAATTA	ACAGGAAATC	ATGCCTTACA	GAGCTTGATA	1380
	TCATCTGAAA	ACTTTCCAGA	ACTCAAGGTT	ATAGAAATGC	CTTATGCTTA	CCAGTGCTGT	1440
20	GCATTTGGAG	TGTGTGAGAA	TGCCTATAAG	ATTTCTAATC	AATGGAATAA	AGGTGACAAC	1500
	AGCAGTATGG	ACGACCTTCA	TAAGAAAGAT	GCTGGAATGT	TTCAGGCTCA	AGATGAACGT	1560
	GACCTTGAAG	ATTTCCTGCT	TGACTTTGAG	GAAGACCTGA	AAGCCCTTCA	TTCAGTGCAG	1620
	TGTTCACCTT	CCCCAGGCCC	CTTCAAACCC	TGTGAACACC	TGCTTGATGG	CTGGCTGATC	1680
	AGAATTGGAG	TGTGGACCAT	AGCAGTTCTG	GCACTTACTT	GTAATGCTTT	GGTGACTTCA	1740
25	ACAGTTTTCA	GATCCCCTCT	GTACATTTCC	CCCATTAAAC	TGTTAATTGG	GGTCATCGCA	1800
	GCAGTGAACA	TGCTCACGGG	AGTCTCCAGT	GCCGTGCTGG	CTGGTGTGGA	TGCGTTCACT	1860
	TTTGGCAGCT	TTGCACGACA	TGGTGCCTGG	TGGGAGAATG	GGGTTGGTTG	CCATGTCATT	1920
	GGTTTTTTGT	CCATTTTTGC	TTCAGAATCA	TCTGTTTTCC	ጥርርጥጥልርጥርጥ	GGCAGCCCTG	1980

	GAGCGTGGGT	TCTCTGTGAA	ATATTCTGCA	AAATTTGAAA	CGAAAGCTCC	ATTTTCTAGC	2040
	CTGAAAGTAA	TCATTTTGCT	CTGTGCCCTG	CTGGCCTTGA	CCATGGCCGC	AGTTCCCCTG	2100
• .	CTGGGTGGCA	GCAAGTATGG	CGCCTCCCCT	CTCTGCCTGC	CTTTGCCTTT	TGGGGAGCCC	2160
	AGCACCATGG	GCTACATGGT	CGCTCTCATC	TTGCTCAATT	CCCTTTGCTT	CCTCATGATG	2220
5	ACCATTGCCT	ACACCAAGCT	CTACTGCAAT	TTGGACAAGG	GAGACCTGGA	GAATATTTGG	2280
	GACTGCTCTA	TGAAAAAACA	CATTGCCCTG	TTGCTCTTCA	CCAACTGCAT	CCTAAACTGC	2340
	CCTGTGGCTT	TCTTGTCCTT	CTCCTCTTTA	ATAAACCTTA	CATTTATCAG	TCCTGAAGTA	2400
	ATTAAGTTTA	TCCTTCTGGT	GGTAGTCCCA	CTTCCTGCAT	GTCTCAATCC	CCTTCTCTAC	2460
	ATCTTGTTCA	ATCCTCACTT	TAAGGAGGAT	CTGGTGAGCC	TGAGAAAGCA	AACCTACGTC	2520
10	TGGACAAGAT	CAAAACACCC	AAGCTTGATG	TCAATTAACT	CTGATGATGT	CGAAAAACAG	2580
	TCCTGTGACT	CAACTCAAGC	CTTGGTAACC	TTTACCAGCT	CCAGCATCAC	TTATGACCTG	2640
	CCTCCCAGTT	CCGTGCCATC	ACCAGCTTAT	CCAGTGACTG	AGAGCTGCCA	TCTTTCCTCT	2700
	GTGGCATTTG	TCCCATGTCT	CTAA				2724
	(279) INFO	RMATION FOR	SEQ ID NO:2	278:	٠		
15	(i) S	SEQUENCE CHA	RACTERISTIC	CS:			

- (A) LENGTH: 907 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- 20 (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:278:

Met Asp Thr Ser Arg Leu Gly Val Leu Leu Ser Leu Pro Val Leu Leu

Gln Leu Ala Thr Gly Gly Ser Ser Pro Arg Ser Gly Val Leu Leu Arg 25. 25

> Gly Cys Pro Thr His Cys His Cys Glu Pro Asp Gly Arg Met Leu Leu 40

> Arg Val Asp Cys Ser Asp Leu Gly Leu Ser Glu Leu Pro Ser Asn Leu . 50 60

30 Ser Val Phe Thr Ser Tyr Leu Asp Leu Ser Met Asn Asn Ile Ser Gln 70

> Leu Leu Pro Asn Pro Leu Pro Ser Leu Arg Phe Leu Glu Glu Leu Arg 85 90

	Leu	Ala	Gly	Asn 100	Ala	Leu	Thr	Tyr	Ile 105	Pro	Lys	Gly	Ala	Phe 110	Thr	Gly
	Leu	Tyr	Ser 115	Leu	Lys	Val	Leu	Met 120	Leu	Gln	Asn	Asn	Gln 125	Leu	Arg	His
5	Val	Pro 130	Thr	Glu	Ala	Leu	Gln 135	Asn	Leu	Arg	Ser	Leu 140	Gln	Ser	Leu	Arg
	Leu 145	Asp	Ala	Asn	His	Ile 150	Ser	Tyr	Val	Pro	Pro 155	Ser	Cys	Phe	Ser	Gly 160
10	Leu	His	Ser	Leu	Arg 165	His	Leu	Trp	Leu	Asp 170	Asp	Asn	Ala	Leu	Thr 175	Glu
	Ile	Pro	Val	Gln 180	Ala	Phe	Arg	Ser	Leu 185	Ser	Ala	Leu	Gln	Ala 190	Met	Thr
	Leu	Ala	Leu 195	Asn	Lys	Ile	His	His 200	Ile	Pro	Asp	Tyr	Ala 205	Phe	Gly	Asn
15	Leu	Ser 210	Ser	Leu	Val	Val	Leu 215	His	Leu	His	Asn	Asn 220	Arg	Ile	His	Ser
	Leu 225	Gly	Lys	Lys	Cys	Phe 230	Asp	Gly	Leu	His	Ser 235	Leu	Glu	Thr	Leu	Asp 240
20					245					250			Ile	_	255	
				260					265				Ile	270		
	Pro	Glu	Lys 275	Ala	Phe	Val	Gly	Asn 280	Pro	Ser	Leu	Ile	Thr 285	Ile	His	Phe
25	Tyr	Asp 290	Asn	Pro	Ile	Gln	Phe 295	Val	Gly	Arg	Ser	Ala 300	Phe	Gln	His	Leu
	Pro 305	Glu	Leu	Arg	Thr	Leu 310	Thr	Leu	Asn	Gly	Ala 315	Ser	Gln	Ile	Thr	Glu 320
30	Phe	Pro	Asp	Leu	Thr 325	Gly	Thr	Ala	Asn	Leu 330	Glu	Ser	Leu	Thr	Leu 335	Thr
	Gly	Ala	Gln	Ile 340	Ser	Ser	Leu	Pro	Gln 345	Thr	Val	Cys	Asn	Gln 350	Leu	Pro
	Asn	Leu	Gln 355	Val	Leu	Asp	Leu	Ser 360	Tyr	Asn	Leu	Leu	Glu 365	Asp	Leu	Pro
35	Ser	Phe 370	Ser	Val	Cys	Gln	Lys 375	Leu	Gln	Lys	Ile	Asp 380	Leu	Arg	His	Asn
	Glu	Ile	Tyr	Glu	Ile	Lys	Val	Asp	Thr	Phe	Gln	Gln	Leu	Leu	Ser	Leu

	385					390	٠.				395	.*				400
	Arg	Ser	Leu	Asn	Leu 405		Trp	Asn	Lys	Ile 410	Ala	Ile	Ile	His	Pro 415	Asn
5	Ala	Phe	Ser	Thr 420	Leu	Pro	Ser	Leu	Ile 425	Lys	Leu	Asp	Leu	Ser 430	Ser	Asn
	Leu	Leu	Ser 435	Ser	Phe	Pro	Ile	Thr 440	Gly	Leu	His	Gly	Leu 445	Thr	His	Leu
	Lys	Leu 450	Thr	Gly	Asn	His	Ala 455	Leu	Gln	Ser	Leu	Ile 460	Ser	Ser	Glu	Asn
10	Phe 465	Pro	Glu	Leu	Lys	Val 470	Ile	Glu	Met	Pro	Tyr 475	Ala	Tyr	Gln	Cys	Cys 480
. N. #*	Ala	Phe	Gly	Val	Cys 485	Glu	Asn	Ala		Lys 490	Ile	Ser	Asn	Gln	Trp 495	Asn
15	Lys	Gly	Asp	Asn 500	Ser	Ser	Met	Asp	Asp 505	Leu	His	Lys	Lys	Asp 510	Ala	Gly
	Met	Phe	Gln 515	Ala	Gln	Asp	Glu	Arg 520	Asp	Leu	Glu	Asp	Phe 525	Leu	Leu	Asp
	Phe	Glu 530	Glu	Asp	Leu	Lys	Ala 535	Leu	His	Ser	Val	Gln 540	Cys	Ser	Pro	Ser
20	Pro 545	Gly	Pro	Phe	Lys	Pro 550	Cys	Glu	His	Leu	Leu 555	Asp	Gly	Trp	Leu	Ile 560
	Arg	Ile	Gly	Val	Trp 565	Thr	Ile	Ala	Val	Leu 570	Ala	Leu	Thr	Cys	Asn 575	Ala
25	Leu	Val	Thr	Ser 580	Thr	Val	Phe	Arg	Ser 585	Pro	Leu	Tyr	Ile	Ser 590	Pro	Ile
	Lys	Leu	Leu 595	Ile	Gly	Val	Ile	Ala 600	Ala	Val	Asn	Met	Leu 605	Thr	Gly	Val
	Ser	Ser 610	Ala	Val	Leu	Ala	Gly 615	Val	Asp	Ala	Phe	Thr 620	Phe	Gly	Ser	Phe
30	Ala 625	Arg	His	Gly	Ala	Trp 630	Trp	Glu	Asn	Gly	Val 635	Gly	Cys	His	Val	Ile 640
	Gly	Phe	Leu	Ser	Ile 645	Phe	Ala	Ser	Glu	Ser 650	Ser	Val	Phe	Leu	Leu 655	Thr
35	Leu	Ala	Ala	Leu 660	Glu	Arg	Gly	Phe	Ser 665	Val	Lys	Tyr	Ser	Ala 670	Lys	Phe
	Glu	Thr	Lys 675	Ala	Pro	Phe	Ser	Ser 680	Leu	Lys	Val	Ile	Ile 685	Leu	Leu	Cys

PCT/US99/23938

235

WO 00/22129

15 Ile Lys Phe Ile Leu Leu Val Val Val Pro Leu Pro Ala Cys Leu 805 Pro Leu Leu Tyr Ile Leu Phe Asn Pro His Phe Lys Glu Asp Leu 820 Ser Leu Arg Lys Gln Thr Tyr Val Trp Thr Arg Ser Lys His Pro 840 Leu Met Ser Ile Asn Ser Asp Asp Val Glu Lys Gln Ser Cys Asp 850 Thr Gln Ala Leu Val Thr Phe Thr Ser Ser Ser Ile Thr Tyr Asp 865 Pro Pro Ser Ser Val Pro Ser Pro Ala Tyr Pro Val Thr Glu Ser 895 His Leu Ser Ser Val Ala Phe Val Pro Cys Leu 905 (280) INFORMATION FOR SEQ ID NO:279:																		
5 Ser Thr Met Gly Tyr Met Val Ala Leu Ile Leu Leu Asn Ser Leu 735 Phe Leu Met Met Thr Ile Ala Tyr Thr Lys Leu Tyr Cys Asn Leu 740 Lys Gly Asp Leu Glu Asn Ile Trp Asp Cys Ser Met Lys Lys His 755 Ala Leu Leu Leu Phe Thr Asn Cys Ile Leu Asn Cys Pro Val Ala 770 Leu Ser Phe Ser Ser Leu Ile Asn Leu Thr Phe Ile Ser Pro Glu 795 Ile Lys Phe Ile Leu Leu Val Val Pro Leu Pro Ala Cys Leu 805 Pro Leu Leu Tyr Ile Leu Phe Asn Pro His Phe Lys Glu Asp Leu 820 Ser Leu Arg Lys Gln Thr Tyr Val Trp Thr Arg Ser Lys His Pro 835 Leu Met Ser Ile Asn Ser Asp Asp Val Glu Lys Gln Ser Cys Asp 855 Thr Gln Ala Leu Val Thr Phe Thr Ser Ser Ser Ile Thr Tyr Asp 865 Pro Pro Ser Ser Val Pro Ser Pro Ala Tyr Pro Val Thr Glu Ser 885 His Leu Ser Ser Val Ala Phe Val Pro Cys Leu 905 (280) INFORMATION FOR SEQ ID NO:279:			Ala		Leu	Ala	Leu	Thr		Ala	Ala	Val	Pro		Leu	Gly	Gly	Ser
Phe Leu Met Met Thr Ile Ala Tyr Thr Lys Leu Tyr Cys Asn Leu Tyr 755 Lys Gly Asp Leu Glu Asn Ile Trp Asp Cys Ser Met Lys Lys His 765 Ala Leu Leu Leu Phe Thr Asn Cys Ile Leu Asn Cys Pro Val Ala 770 Leu Ser Phe Ser Ser Leu Ile Asn Leu Thr Phe Ile Ser Pro Glu 785 Ile Lys Phe Ile Leu Leu Val Val Val Pro Leu Pro Ala Cys Leu 815 Pro Leu Leu Tyr Ile Leu Phe Asn Pro His Phe Lys Glu Asp Leu 820 Ser Leu Arg Lys Gln Thr Tyr Val Trp Thr Arg Ser Lys His Pro 835 Leu Met Ser Ile Asn Ser Asp Asp Val Glu Lys Gln Ser Cys Asp 855 Thr Gln Ala Leu Val Thr Phe Thr Ser Ser Ser Ile Thr Tyr Asp 865 Pro Pro Ser Ser Val Ala Phe Val Pro Cys Leu 905 His Leu Ser Ser Val Ala Phe Val Pro Cys Leu 905 (280) INFORMATION FOR SEQ ID NO:279:				Tyr	Gly	Ala	Ser		Leu	Cys	Leu	Pro		Pro	Phe	Gly	Glu	Pro 720
10 Lys Gly Asp Leu Glu Asn Ile Trp Asp Cys Ser Met Lys Lys His 765 Ala Leu Leu Leu Phe Thr Asn Cys Ile Leu Asn Cys Pro Val Ala 770 Leu Ser Phe Ser Ser Leu Ile Asn Leu Thr Phe Ile Ser Pro Glu 795 15 Ile Lys Phe Ile Leu Leu Val Val Val Pro Leu Pro Ala Cys Leu 815 Pro Leu Leu Tyr Ile Leu Phe Asn Pro His Phe Lys Glu Asp Leu 820 Ser Leu Arg Lys Gln Thr Tyr Val Trp Thr Arg Ser Lys His Pro 835 Leu Met Ser Ile Asn Ser Asp Asp Val Glu Lys Gln Ser Cys Asp 855 Thr Gln Ala Leu Val Thr Phe Thr Ser Ser Ser Ile Thr Tyr Asp 865 Pro Pro Ser Ser Val Pro Ser Pro Ala Tyr Pro Val Thr Glu Ser 895 His Leu Ser Ser Val Ala Phe Val Pro Cys Leu 905 (280) INFORMATION FOR SEQ ID NO:279:	5		Ser	Thr	Met	Gly		Met	Val	Ala	Leu		Leu	Leu	Asn	Ser		Cys
10 755 760 765 Ala Leu Leu Leu Phe Thr Asn Cys Ile Leu Asn Cys Pro Val Ala 770 Leu Ser Phe Ser Ser Leu Ile Asn Leu Thr Phe Ile Ser Pro Glu 785 15 11e Lys Phe Ile Leu Leu Val Val Val Pro Leu Pro Ala Cys Leu 805 Pro Leu Leu Tyr Ile Leu Phe Asn Pro His Phe Lys Glu Asp Leu 820 Ser Leu Arg Lys Gln Thr Tyr Val Trp Thr Arg Ser Lys His Pro 835 Leu Met Ser Ile Asn Ser Asp Asp Val Glu Lys Gln Ser Cys Asp 850 Thr Gln Ala Leu Val Thr Phe Thr Ser Ser Ser Ile Thr Tyr Asp 865 Pro Pro Ser Ser Val Ala Pro Ser Pro Ala Tyr Pro Val Thr Glu Ser 895 His Leu Ser Ser Val Ala Phe Val Pro Cys Leu 905 (280) INFORMATION FOR SEQ ID NO:279:			Phe	Leu	Met		Thr	Ile	Ala	Tyr		Lys	Leu	Tyr	Cys		Leu	Asp
Leu Ser Phe Ser Ser Leu Ile Asn Leu Thr Phe Ile Ser Pro Glu 785 Ile Lys Phe Ile Leu Leu Val Val Pro Leu Pro Ala Cys Leu 805 Pro Leu Leu Tyr Ile Leu Phe Asn Pro His Phe Lys Glu Asp Leu 820 Ser Leu Arg Lys Gln Thr Tyr Val Trp Thr Arg Ser Lys His Pro 835 Leu Met Ser Ile Asn Ser Asp Asp Val Glu Lys Gln Ser Cys Asp 850 Thr Gln Ala Leu Val Thr Phe Thr Ser Ser Ser Ile Thr Tyr Asp 865 Pro Pro Ser Ser Val Pro Ser Pro Ala Tyr Pro Val Thr Glu Ser 885 His Leu Ser Ser Val Ala Phe Val Pro Cys Leu 900 (280) INFORMATION FOR SEQ ID NO:279:	10		Lys	Gly		Leu	Glu	Asn	Ile		Asp	Cys	Ser	Met		Lys	His	Ile
15			Ala		Leu	Leu	Phe	Thr		Cys	Ile	Leu	Asn		Pro	Val	Ala	Phe
805 810 815 Pro Leu Leu Tyr Ile Leu Phe Asn Pro His Phe Lys Glu Asp Leu 820 Ser Leu Arg Lys Gln Thr Tyr Val Trp Thr Arg Ser Lys His Pro 845 Leu Met Ser Ile Asn Ser Asp Asp Val Glu Lys Gln Ser Cys Asp 850 Thr Gln Ala Leu Val Thr Phe Thr Ser Ser Ser Ile Thr Tyr Asp 865 Pro Pro Ser Ser Val Pro Ser Pro Ala Tyr Pro Val Thr Glu Ser 885 His Leu Ser Ser Val Ala Phe Val Pro Cys Leu 905 (280) INFORMATION FOR SEQ ID NO:279:				Ser	Phe	Ser	Ser		Ile	Asn	Leu	Thr		Ile	Ser	Pro	Glu	Val 800
Ser Leu Arg Lys Gln Thr Tyr Val Trp Thr Arg Ser Lys His Pro 835 Leu Met Ser Ile Asn Ser Asp Asp Val Glu Lys Gln Ser Cys Asp 850 Thr Gln Ala Leu Val Thr Phe Thr Ser Ser Ser Ile Thr Tyr Asp 865 Pro Pro Ser Ser Val Pro Ser Pro Ala Tyr Pro Val Thr Glu Ser 885 His Leu Ser Ser Val Ala Phe Val Pro Cys Leu 900 (280) INFORMATION FOR SEQ ID NO:279:	15	,	Ile	Lys	Phe	Ile		Leu	Val	Val	Val		Leu	Pro	Ala	Cys		Asn
Leu Met Ser Ile Asn Ser Asp Asp Val Glu Lys Gln Ser Cys Asp 850 Thr Gln Ala Leu Val Thr Phe Thr Ser Ser Ser Ile Thr Tyr Asp 865 Pro Pro Ser Ser Val Pro Ser Pro Ala Tyr Pro Val Thr Glu Ser 890 His Leu Ser Ser Val Ala Phe Val Pro Cys Leu 900 (280) INFORMATION FOR SEQ ID NO:279:			Pro	Leu	Leu		Ile	Leu	Phe	Asn		His	Phe	Lys	Glu	_	Leu	Val
Thr Gln Ala Leu Val Thr Phe Thr Ser Ser Ser Ile Thr Tyr Asp 865 Pro Pro Ser Ser Val Pro Ser Pro Ala Tyr Pro Val Thr Glu Ser 885 His Leu Ser Ser Val Ala Phe Val Pro Cys Leu 900 (280) INFORMATION FOR SEQ ID NO:279:	20		Ser	Leu		Lys	Gln	Thr	Tyr		Trp	Thr	Arg	Ser		His	Pro	Ser
Pro Pro Ser Ser Val Pro Ser Pro Ala Tyr Pro Val Thr Glu Ser 885 His Leu Ser Ser Val Ala Phe Val Pro Cys Leu 900 (280) INFORMATION FOR SEQ ID NO:279:			Leu		Ser	Ile	Asn	Ser		Asp	Val	Glu	Lys		Ser	Cys	qaA	Ser
His Leu Ser Ser Val Ala Phe Val Pro Cys Leu 900 905 (280) INFORMATION FOR SEQ ID NO:279:		•		Gln	Ala	Leu	Val		Phe	Thr	Ser	Ser		Ile	Thr	Tyr	Asp	Leu 880
900 905 (280) INFORMATION FOR SEQ ID NO:279:	25		Pro	Pro	Ser	Ser		Pro	Ser	Pro	Ala		Pro	Val	Thr	Glu		Cys
			His	Leu	Ser		Val	Ala	Phe	Val		Cys	Leu					
		(280)	INE	ORMA	OITA	FOF	SEÇ) ID	NO:2	79:								
30 (i) SEQUENCE CHARACTERISTICS:	30		(i)	SEÇ	UENC	E CH	IARAC	TERI	STIC	S:								

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- 35 (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:279:

CATGCCAACC GGCCCGCGAG GCTGCTGCTG GT

(281) INFORMATION FOR SEQ ID NO:280:

5

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 32 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:280:

ACCAGCAGCA GCCTCGCGGG CCGGTTGGCA TG

32

RNSDOCID- - WO ... noostock + 1 --

Int. Ional Application No PCT/US 99/23938

		1 01/03 99/23938
A. CLASSI IPC 7	FICATION OF SUBJECT MATTER C12N15/12 C07K14/72 G01N33	3/50 G01N33/566
According to	o International Patent Classification (IPC) or to both national class	iffication and IPC
B. FIELDS	SEARCHED	
Minimum do IPC 7	cumentation searched (classification system tollowed by classific C12N C07K G01N	cation symbols)
Documental	tion searched other than minimum documentation to the extent the	at such documents are included in the fields searched
Electronic d	ata base consulted during the international search (name of data	base and, where practical, search terms used)
C. DOCUME	ENTS CONSIDERED TO BE RELEVANT	
Category °	Citation of document, with indication, where appropriate, of the	relevant passages Relevant to claim No.
X	KJELSBERG M. A. ET AL.: "CONST ACTIVATION OF THE ALPHA1B-ADREN RECEPTOR BY ALL AMINO ACID SUBS AT A SINGLE SITE" JOURNAL OF BIOLOGICAL CHEMISTRY vol. 267, no. 3, 25 January 1992 (1992-01-25), pa 1430-1433, XP002911764 ISSN: 0021-9258 the whole document	ERGIC 4-13, TITUTIONS 15-33, 35-37,41
X Furthe	er documents are listed in the continuation of box C.	X Patent family members are listed in annex.
'A" documer conside earlier do filling da documen which is citation O" documen other m P" documen later the	at which may throw doubts on priority claim(s) or s cited to establish the publication date of another or other special reason (as specified) nt referring to an oral disclosure, use, exhibition or	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "&" document member of the same patent family
	March 2000	Date of mailing of the international search report 09/03/2000
lame and ma	ailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Mand 1 , B

C.(Continue Category	number Comments Constitution		/	
Jalegory 1	ation) DOCUMENTS CONSIDERED TO BE RELEVANT			
	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.		
X .	SCHEER A. ET AL.: "CONSTITUTIVELY ACTIVE G PROTEIN-COUPLED RECEPTORS: POTENTIAL MECHANISMS OF RECEPTOR ACTIVATION" JOURNAL OF RECEPTOR AND SIGNAL TRANSDUCTION RESEARCH, vol. 17, no. 1/03, 1997, pages 57-73, XP000867531 ISSN: 1079-9893 the whole document		1,2, 4-13, 15-33, 35-37,41	
(WO 97 21731 A (NEW ENGLAND MEDICAL CENTER INC) 19 June 1997 (1997-06-19)		1,2,4, 9-13, 20-32, 35-37,41	
	the whole document, especially Fig. 2-3			
	WO 98 38217 A (HERRICK DAVIS KATHARINE ;TEITLER MILT (US); EGAN CHRISTINA C (US)) 3 September 1998 (1998-09-03)		1,2, 4-13, 15-33,	
	the whole document, especially page 7, lines 24-27, and figure 4		35-37,41	
, X	PAUWELS P. J. ET AL.: "REVIEW: AMINO ACID DOMAINS INVOLVED IN CONSTITUTIVE ACTIVATION OF G-PROTEIN-COUPLED RECEPTORS" MOLECULAR NEUROBIOLOGY, vol. 17, no. 1/03, 1998, pages 109-135, XP000866477 ISSN: 0893-7648 the whole document		1,2, 4-13, 15-33, 35-37,41	
1				

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

. _rnational application No.

Line Court Court Control

PCT/US 99/23938

Box I	Observations whire certain claims were found unsearchable (Continuation of item 1 of first she_t)
This Inte	rnational Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:
1.	Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
2. X	Claims Nos.: 34,38-40,42,43 because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically: See FURTHER INFORMATION sheet PCT/ISA/210
3.	Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).
Box II	Observations where unity of invention is lacking (Continuation of item 2 of first sheet)
This Inte	rnational Searching Authority found multiple inventions in this international application, as follows:
1.	As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2.	As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3.	As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4.	No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
Remark (The additional search fees were accompanied by the applicant's protest. No protest accompanied the payment of additional search fees.

International Application No. PCT/US 99 23938

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 34,38-40,42,43

Claims 34, 38-40, 42 and 43 refer to compounds with an agonistic effect on a GPCR without giving a true technical characterization. Moreover, no such specific compounds are defined in the application. In consequence, the scope of said claims is ambiguous and vague, and their subject-matter is not sufficiently disclosed and supported (Art. 5 and 6 PCT). No search can be carried out for such purely speculative claims whose wording is, in fact, a mere recitation of the results to be achieved.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

information on patent family members

Int. ionel Application No
PCT/US 99/23938

Patent document cited in search repor	t	Publication date		atent family member(s)	Publication dat
WO 9721731	Α	19-06-1997	US	5750353 A	12-05-1998
			AU	1334397 A	03-07-1997
•			CA	2239293 A	19-06-1997
•			EP	0869975 A	14-10-1998
WO 9838217	Α	03-09-1998	AU	6343998 A	18-09-1998

.



PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7:
C12N 15/12, C07K 14/72, G01N 33/50, 33/566

A1

(11) International Publication Number:

WO 00/22129

(43) International Publication Date:

20 April 2000 (20.04,00)

(21) International Application Number:

PCT/US99/23938

(22) International Filing Date:

12 October 1999 (12.10.99)

(30) Priority Data:

09/170,496

13 October 1998 (13.10.98)

US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application

US

09/170,496 (CIP)

Filed on

13 October 1998 (13.10.98)

(71) Applicant (for all designated States except US): ARENA PHARMACEUTICALS, INC. [US/US]; 6166 Nancy Ridge Drive, San Diego, CA 92121 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): BEHAN, Dominic, P. [GB/US]; 11472 Roxboro Court, San Diego, CA 92131 (US). CHALMERS, Derek, T. [GB/US]; 347 Longden Lane, Solana Beach, CA 92075 (US). L1AW, Chen, W. [US/US]; 7668 Salix Place, San Diego, CA 92129 (US).

(74) Agents: MILLER, Suzanne, E. et al.; Woodcock Washburn Kurtz Mackiewicz & Norris LLP, 46th floor, One Liberty Place, Philadelphia, PA 19103 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: NON-ENDOGENOUS, CONSTITUTIVELY ACTIVATED HUMAN G PROTEIN-COUPLED RECEPTORS

(57) Abstract

Disclosed herein are constitutively activated, non-endogenous versions of endogenous human G protein-coupled receptors comprising (a) the following amino acid sequence region (C-terminus to N-terminus orientation) and/or (b) the following nucleic acid sequence region (3' to 5' orientation) transversing the transmembrane-6 (TM6) and intracellular loop-3 (IC3) regions of the GPCR: (a) P¹ AA15 X and/or (b) Pcodon (AA-codon)15 Xcodon, respectively. In a most preferred embodiment, P¹ and Pcodon are endogenous proline and an endogenous nucleic acid encoding region encoding proline, respectively, located within TM6 of the non-endogenous GPCR; AA15 and (AA-codon)15 are 15 endogenous amino acid residues and 15 codons encoding endogenous amino acid residues, respectively; and X and Xcodon are non-endogenous lysine and a non-endogenous nucleic acid encoding region encoding lysine, respectively, located within IC3 of the non-endogenous GPCR. Because it is most preferred that the non-endogenous human GPCRs which incorporate these mutations are incorporated into mammalian cells and utilized for the screening of the candidate compounds, the non-endogenous human GPCR incorporating the mutation need not be purified and isolated *per se* (i.e., these are incorporated within the cellular membrane of a mammalian cell), although such purified and isolated non-endogenous human GPCRs are well within the purview of this disclosure.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	Fl	Finland	LT	Lithuania	SK	Slovakia
AΤ	Austria	FR	France	LU	Luxembourg	SN	Senegal
ΑU	Australia	GA ·	Gabon	LV	Latvia	SZ	Swaziland
ΑZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of Americ
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	zw	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	Li	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

WO 00/22129 PCT/US99/23938

NON-ENDOGENOUS, CONSTITUTIVELY ACTIVATED HUMAN G PROTEIN-COUPLED RECEPTORS

The benefits of commonly owned U.S. Serial Number 09/170,496, filed October 13, 1998, U.S. Serial Number 08/839, 449 filed April 14, 1997 (now abandoned), U.S. Serial Number 09/060,188, filed April 14, 1998; U.S. Provisional Number 60/090,783, filed June 26, 1998; and U.S. Provisional Number 60/095,677, filed on August 7, 1998, are hereby claimed. Each of the foregoing applications are incorporated by reference herein in their entirety.

FIELD OF THE INVENTION

The invention disclosed in this patent document relates to transmembrane receptors, and more particularly to human G protein-coupled receptors (GPCRs) which have been altered such that altered GPCRs are constitutively activated. Most preferably, the altered human GPCRs are used for the screening of therapeutic compounds.

BACKGROUND OF THE INVENTION

Although a number of receptor classes exist in humans, by far the most abundant and therapeutically relevant is represented by the G protein-coupled receptor (GPCR or GPCRs) class. It is estimated that there are some 100,000 genes within the human genome, and of these, approximately 2% or 2,000 genes, are estimated to code for GPCRs. Of these, there are approximately 100 GPCRs for which the endogenous ligand that binds to the GPCR has been identified. Because of the significant time-lag that exists between the discovery of an endogenous GPCR and its endogenous ligand, it can be presumed that the remaining 1,900 GPCRs will be identified and characterized long before the endogenous ligands for these receptors are identified. Indeed, the rapidity by which the Human Genome Project is sequencing the 100,000 human genes indicates that the remaining human GPCRs will be fully sequenced within the next few years. Nevertheless, and despite the efforts to sequence the human genome, it is still very unclear as to how scientists will be able to rapidly, effectively and efficiently exploit this information to improve and enhance the human condition. The present invention is geared towards this important objective.

10

15

Receptors, including GPCRs, for which the endogenous ligand has been identified are referred to as "known" receptors, while receptors for which the endogenous ligand has not been identified are referred to as "orphan" receptors. This distinction is not merely semantic, particularly in the case of GPCRs. GPCRs represent an important area for the development of pharmaceutical products: from approximately 20 of the 100 known GPCRs, 60% of all prescription pharmaceuticals have been developed. Thus, the orphan GPCRs are to the pharmaceutical industry what gold was to California in the late 19th century – an opportunity to drive growth, expansion, enhancement and development. A serious drawback exists, however,

WO 00/22129 PCT/US99/23938

- 3 -

5

10

15

with orphan receptors relative to the discovery of novel therapeutics. This is because the traditional approach to the discovery and development of pharmaceuticals has required access to both the receptor *and* its endogenous ligand. Thus, heretofore, orphan GPCRs have presented the art with a tantalizing and undeveloped resource for the discovery of pharmaceuticals.

Under the traditional approach to the discovery of potential therapeutics, it is generally the case that the receptor is first identified. Before drug discovery efforts can be initiated, elaborate, time consuming and expensive procedures are typically put into place in order to identify, isolate and generate the receptor's endogenous ligand - this process can require from between 3 and ten years per receptor, at a cost of about \$5million (U.S.) per receptor. These time and financial resources must be expended before the traditional approach to drug discovery can commence. This is because traditional drug discovery techniques rely upon so-called "competitive binding assays" whereby putative therapeutic agents are "screened" against the receptor in an effort to discover compounds that either block the endogenous ligand from binding to the receptor ("antagonists"), or enhance or mimic the effects of the ligand binding to the receptor ("agonists"). The overall objective is to identify compounds that prevent cellular activation when the ligand binds to the receptor (the antagonists), or that enhance or increase cellular activity that would otherwise occur if the ligand was properly binding with the receptor (the agonists). Because the endogenous ligands for orphan GPCRs are by definition not identified, the ability to discover novel and unique therapeutics to these receptors using traditional drug discovery techniques is not possible. The present invention, as will be set forth in greater detail below, overcomes these and other severe limitations created by such traditional drug discovery techniques.

GPCRs share a common structural motif. All these receptors have seven sequences of between 22 to 24 hydrophobic amino acids that form seven alpha helices, each of which spans the

membrane (each span is identified by number, *i.e.*, transmembrane-1 (TM-1), transmebrane-2 (TM-2), etc.). The transmembrane helices are joined by strands of amino acids between transmembrane-2 and transmembrane-3, transmembrane-4 and transmembrane-5, and transmembrane-6 and transmembrane-7 on the exterior, or "extracellular" side, of the cell membrane (these are referred to as "extracellular" regions 1, 2 and 3 (EC-1, EC-2 and EC-3), respectively). The transmembrane helices are also joined by strands of amino acids between transmembrane-1 and transmembrane-2, transmembrane-3 and transmembrane-4, and transmembrane-5 and transmembrane-6 on the interior, or "intracellular" side, of the cell membrane (these are referred to as "intracellular" regions 1, 2 and 3 (IC-1, IC-2 and IC-3), respectively). The "carboxy" ("C") terminus of the receptor lies in the intracellular space within the cell, and the "amino" ("N") terminus of the receptor lies in the extracellular space outside of the cell. The general structure of G protein-coupled receptors is depicted in Figure 1.

Generally, when an endogenous ligand binds with the receptor (often referred to as "activation" of the receptor), there is a change in the conformation of the intracellular region that allows for coupling between the intracellular region and an intracellular "G-protein." Although other G proteins exist, currently, Gq, Gs, Gi, and Go are G proteins that have been identified. Endogenous ligand-activated GPCR coupling with the G-protein begins a signaling cascade process (referred to as "signal transduction"). Under normal conditions, signal transduction ultimately results in cellular activation or cellular inhibition. It is thought that the IC-3 loop as well as the carboxy terminus of the receptor interact with the G protein. A principal focus of this invention is directed to the transmembrane-6 (TM6) region and the intracellular-3 (IC3) region of the GPCR.

Under physiological conditions, GPCRs exist in the cell membrane in equilibrium between

WO 00/22129 PCT/US99/23938

- 5 -

5

15

20

two different conformations: an "inactive" state and an "active" state. As shown schematically in Figure 2, a receptor in an inactive state is unable to link to the intracellular signaling transduction pathway to produce a biological response. Changing the receptor conformation to the active state allows linkage to the transduction pathway (via the G-protein) and produces a biological response.

A receptor may be stabilized in an active state by an endogenous ligand or a compound such as a drug. Recent discoveries, including but not exclusively limited to modifications to the amino acid sequence of the receptor, provide means other than endogenous ligands or drugs to promote and stabilize the receptor in the active state conformation. These means effectively stabilize the receptor in an active state by simulating the effect of an endogenous ligand binding to the receptor. Stabilization by such ligand-independent means is termed "constitutive receptor activation."

As noted above, the use of an orphan receptor for screening purposes has not been possible. This is because the traditional "dogma" regarding screening of compounds mandates that the ligand for the receptor be known. By definition, then, this approach has no applicability with respect to orphan receptors. Thus, by adhering to this dogmatic approach to the discovery of therapeutics, the art, in essence, has taught and has been taught to forsake the use of orphan receptors unless and until the endogenous ligand for the receptor is discovered. Given that there are an estimated 2,000 G protein coupled receptors, the majority of which are orphan receptors, such dogma castigates a creative, unique and distinct approach to the discovery of therapeutics.

Information regarding the nucleic acid and/or amino acid sequences of a variety of GPCRs is summarized below in Table A. Because an important focus of the invention disclosed herein is directed towards orphan GPCRs, many of the below-cited references are related to orphan GPCRs. However, this list is not intended to imply, nor is this list to be construed, legally or

otherwise, that the invention disclosed herein is only applicable to orphan GPCRs or the specific GPCRs listed below. Additionally, certain receptors that have been isolated are not the subject of publications per se; for example, reference is made to a G Protein-Coupled Receptor database on the "world-wide web" (neither the named inventors nor the assignee have any affiliation with this site) that lists GPCRs. Other GPCRs are the subject of patent applications owned by the present assignce and these are not listed below (including GPR3, GPR6 and GPR12; see U.S. Provisional Number 60/094879):

Table A

	Receptor Name	Publication Reference
10	GPR1	23 Genomics 609 (1994)
	GPR4	14 DNA and Cell Biology 25 (1995)
	GPR5	14 DNA and Cell Biology 25 (1995)
	GPR7	28 Genomics 84 (1995)
	GPR8	28 Genomics 84 (1995)
15	GPR9	184 J. Exp. Med. 963 (1996)
	GPR10	29 Genomics 335 (1995)
	GPR15	32 Genomics 462 (1996)
4. A	GPR17	70 J Neurochem. 1357 (1998)
	GPR18	42 Genomics 462 (1997)
20	GPR20	187 Gene 75 (1997)
	GPR21	187 Gene 75 (1997)
	GPR22	187 Gene 75 (1997)
••	GPR24	398 FEBS Lett. 253 (1996)
	GPR30	45 Genomics 607 (1997)
25	GPR31	42 Genomics 519 (1997)
	GPR32	50 Genomics 281 (1997)
	GPR40	239 Biochem. Biophys.
		Res. Commun. 543 (1997)
	GPR41	239 Biochem. Biophys.
		Res. Commun. 543 (1997)
	GPR43	239 Biochem. Biophys.
		Res. Commun. 543 (1997)
30	APJ	136 Gene 355 (1993)
	BLR1	22 Eur. J. Immunol. 2759 (1992)
	CEPR	231 Biochem. Biophys.
		Res. Commun. 651 (1997)
	EBI1	23 Genomics 643 (1994)
	EBI2	67 J. Virol. 2209 (1993)
35	ETBR-LP2	424 FEBS Lett. 193 (1998)
	GPCR-CNS	54 Brain Res. Mol. Brain Res. 152 (1998);
		45 Genomics 68 (1997)
	GPR-NGA	394 FEBS Lett. 325 (1996)
	H9	386 FEBS Lett 219 (1996)

WO 00/22129 PCT/US99/23938

-7-

HBA954	1261 Biochim. Biophys. Acta 121 (1995)	
HG38	247 Biochem. Biophys.	
	Res. Commun. 266 (1998)	
HM74	5 Int. Immunol. 1239 (1993)	
OGR1	35 Genomics 397 (1996)	
V28	163 Gene 295 (1995)	

As will be set forth and disclosed in greater detail below, utilization of a mutational cassette to modify the endogenous sequence of a human GPCR leads to a constitutively activated version of the human GPCR. These non-endogenous, constitutively activated versions of human GPCRs can be utilized, *inter alia*, for the screening of candidate compounds to directly identify compounds of, *e.g.*, therapeutic relevance.

SUMMARY OF THE INVENTION

Disclosed herein is a non-endogenous, human G protein-coupled receptor comprising (a) as a most preferred amino acid sequence region (C-terminus to N-terminus orientation) and/or (b) as a most preferred nucleic acid sequence region (3' to 5' orientation) transversing the transmembrane-6 (TM6) and intracellular loop-3 (IC3) regions of the GPCR:

wherein:

15

20

- (1) P¹ is an amino acid residue located within the TM6 region of the GPCR, where P¹ is selected from the group consisting of (i) the endogenous GPCR's proline residue, and (ii) a non-endogenous amino acid residue other than proline;
- (2) AA₁₅ are 15 amino acids selected from the group consisting of

(a) the endogenous GPCR's amino acids (b) non-endogenous amino acid residues, and (c) a combination of the endogenous GPCR's amino acids and non-endogenous amino acids, excepting that none of the 15 endogenous amino acid residues that are positioned within the TM6 region of the GPCR is proline; and

(3) X is a non-endogenous amino acid residue located within the IC3 region of said GPCR, preferably selected from the group consisting of lysine, hisitidine and arginine, and most preferably lysine, excepting that when the endogenous amino acid at position X is lysine, then X is an amino acid other than lysine, preferably alanine;

and/or

5

10

(b) P^{codon} (AA-codon)₁₅ X_{codon}

(1)

(2)

15 wherein:

P^{codon} is a nucleic acid sequence within the TM6 region of the GPCR, where P^{codon} encodes an amino acid selected from the group consisting of (i) the endogenous GPCR's proline residue, and (ii) a non-endogenous amino acid residue other than proline; (AA-codon)₁₅ are 15 codons encoding 15 amino acids selected from the group consisting of (a) the endogenous GPCR's amino acids (b) non-endogenous amino acid residues and (c) a

combination of the endogenous GPCR's amino acids and non-

20

באפרורות בואח המסיים באו

 \cdot (3)

endogenous amino acids, excepting that none of the 15 endogenous codons within the TM6 region of the GPCR encodes a proline amino acid residue; and

X_{codon} is a nucleic acid encoding region residue located within the IC3 region of said GPCR, where X_{codon} encodes a non-endogenous amino acid, preferably selected from the group consisting of lysine, hisitidine and arginine, and most preferably lysine, excepting that when the endogenous encoding region at position X_{codon} encodes the amino acid lysine, then X_{codon} encodes an amino acid other than lysine, preferably alanine.

10

15

The terms endogenous and non-endogenous in reference to these sequence cassettes are relative to the endogenous GPCR. For example, once the endogenous proline residue is located within the TM6 region of a particular GPCR, and the 16th amino acid therefrom is identified for mutation to constitutively activate the receptor, it is also possible to mutate the endogenous proline residue (*i.e.*, once the marker is located and the 16th amino acid to be mutated is identified, one may mutate the marker itself), although it is most preferred that the proline residue not be mutated. Similarly, and while it is most preferred that AA₁₅ be maintained in their endogenous forms, these amino acids may also be mutated. The only amino acid that must be mutated in the non-endogenous version of the human GPCR is X *i.e.*, the endogenous amino acid that is 16 residues from P¹ cannot be maintained in its endogenous form and must be mutated, as further disclosed herein. Stated again, while it is preferred that in the non-endogenous version of the human GPCR, P¹ and AA₁₅ remain in their endogenous forms (*i.e.*, identical to their wild-type forms), once X is identified and mutated, any and/or all of P¹ and AA₁₅ can be mutated. This applies to the nucleic

acid sequences as well. In those cases where the endogenous amino acid at position X is lysine, then in the non-endogenous version of such GPCR, X is an amino acid other than lysine, preferably alanine.

Accordingly, and as a hypothetical example, if the endogenous GPCR has the following endogenous amino acid sequence at the above-noted positions:

P-AACCTTGGRRRDDDE -O

then any of the following exemplary and hypothetical cassettes would fall within the scope of the disclosure (non-endogenous amino acids are set forth in bold):

P-AACCTTGGRRRDDDE -K

10

15

5

P-AACCTTHIGRRDDDE -K

P-ADEETTGGRRRDDDE -A

P-LLKFMSTWZLVAAPO -K

A-LLKFMSTWZLVAAPQ -K

It is also possible to add amino acid residues within AA₁₅, but such an approach is not particularly advanced. Indeed, in the most preferred embodiments, the only amino acid that differs in the non-endogenous version of the human GPCR as compared with the endogenous version of that GPCR is the amino acid in position X; mutation of this amino acid itself leads to constitutive activation of the receptor.

Thus, in particularly preferred embodiments, P¹ and Pcodon are endogenous proline and an endogenous nucleic acid encoding region encoding proline, respectively; and X and X_{codon} are non-endogenous lysine or alanine and a non-endogenous nucleic acid encoding region encoding lysine or alanine, respectively, with lysine being most preferred. Because it is most preferred that the non-endogenous versions of the human GPCRs which incorporate these mutations are

15

incorporated into mammalian cells and utilized for the screening of candidate compounds, the non-endogenous human GPCR incorporating the mutation need not be purified and isolated *per se* (*i.e.*, these are incorporated within the cellular membrane of a mammalian cell), although such purified and isolated non-endogenous human GPCRs are well within the purview of this disclosure. Genetargeted and transgenic non-human mammals (preferably rats and mice) incorporating the non-endogenous human GPCRs are also within the purview of this invention; in particular, genetargeted mammals are most preferred in that these animals will incorporate the non-endogenous versions of the human GPCRs in place of the non-human mammal's endogenous GPCR-encoding region (techniques for generating such non-human mammals to replace the non-human mammal's protein encoding region with a human encoding region are well known; see, for example, U.S. Patent No. 5,777,194.)

It has been discovered that these changes to an endogenous human GPCR render the GPCR constitutively active such that, as will be further disclosed herein, the non-endogenous, constitutively activated version of the human GPCR can be utilized for, *inter alia*, the direct screening of candidate compounds without the need for the endogenous ligand. Thus, methods for using these materials, and products identified by these methods are also within the purview of the following disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a generalized structure of a G protein-coupled receptor with the numbers

20 assigned to the transmembrane helixes, the intracellular loops, and the extracellular loops.

Figure 2 schematically shows the two states, active and inactive, for a typical G protein coupled receptor and the linkage of the active state to the second messenger transduction pathway.

Figure 3 is a sequence diagram of the preferred vector pCMV, including restriction enzymen site locations.

Figure 4 is a diagrammatic representation of the signal measured comparing pCMV, non-endogenous, constitutively active GPR30 inhibition of GPR6-mediated activation of CRE-Luc reporter with endogenous GPR30 inhibition of GPR6-mediated activation of CRE-Luc reporter.

Figure 5 is a diagrammatic representation of the signal measured comparing pCMV, non-endogenous, constitutively activated GPR17 inhibition of GPR3-mediated activation of CRE-Luc reporter with endogenous GPR17 inhibition of GPR3-mediated activation of CRE-Luc reporter.

Figure 6 provides diagrammatic results of the signal measured comparing control pCMV, endogenous APJ and non-endogenous APJ.

Figure 7 provides an illustration of $\rm IP_3$ production from non-endogenous human 5-HT_{2A} receptor as compared to the endogenous version of this receptor.

Figure 8 are dot-blot format results for GPR1 (8A), GPR30 (8B) and APJ (8C).

DETAILED DESCRIPTION

The scientific literature that has evolved around receptors has adopted a number of terms to refer to ligands having various effects on receptors. For clarity and consistency, the following definitions will be used throughout this patent document. To the extent that these definitions conflict with other definitions for these terms, the following definitions shall control:

AGONISTS shall mean compounds that activate the intracellular response when they bind to the receptor, or enhance GTP binding to membranes.

25

30

- 13 -

AMINO ACID ABBREVIATIONS used herein are set below:

	ALANINE	ALA	Α
	ARGININE	ARG	R
1 - 8 °	ASPARAGINE	ASN	N
5 .	ASPARTIC ACID	ASP	D.
	CYSTEINE	CYS	С
	GLUTAMIC ACID	GLU	E
	GLUTAMINE	GLN	Q
	GLYCINE	GLY	G
10	HISTIDINE	HIS	Н
	ISOLEUCINE	ILE	I
	LEUCINE	LEU	L.
	LYSINE	LYS	K
	METHIONINE	MET	М
15	PHENYLALANINE	PHE	F
	PROLINE	PRO	. P
	SERINE	SER	S
, ,	THREONINE	THR	Т
	TRYPTOPHAN	TRP	w
20	TYROSINE	TYR	Y
··.	VALINE	VAL	v · ·

PARTIAL AGONISTS shall mean compounds which activate the intracellular response when they bind to the receptor to a lesser degree/extent than do agonists, or enhance GTP binding to membranes to a lesser degree/extent than do agonists

ANTAGONIST shall mean compounds that competitively bind to the receptor at the same site as the agonists but which do not activate the intracellular response initiated by the active form of the receptor, and can thereby inhibit the intracellular responses by agonists or partial agonists. ANTAGONISTS do not diminish the baseline intracellular response in the absence of an agonist or partial agonist.

CANDIDATE COMPOUND shall mean a molecule (for example, and not limitation, a chemical compound) which is amenable to a screening technique. Preferably, the phrase

"candidate compound" does not include compounds which were publicly known to be compounds selected from the group consisting of inverse agonist, agonist or antagonist to a receptor, as previously determined by an indirect identification process ("indirectly identified compound"); more preferably, not including an indirectly identified compound which has previously been determined to have therapeutic efficacy in at least one mammal; and, most preferably, not including an indirectly identified compound which has previously been determined to have therapeutic utility in humans.

CODON shall mean a grouping of three nucleotides (or equivalents to nucleotides) which generally comprise a nucleoside (adenosine (A), guanosine (G), cytidine (C), uridine (U) and thymidine (T)) coupled to a phosphate group and which, when translated, encodes an amino acid.

COMPOUND EFFICACY shall mean a measurement of the ability of a compound to

inhibit or stimulate receptor functionality, as opposed to receptor binding affinity. A preferred means of detecting compound efficacy is via measurement of, e.g., [35S]GTPyS binding, as further disclosed in the Example section of this patent document.

CONSTITUTIVELY ACTIVATED RECEPTOR shall mean a receptor subject to constitutive receptor activation. In accordance with the invention disclosed herein, a non-endogenous, human constitutively activated G protein-coupled receptor is one that has been mutated to include the amino acid cassette P¹AA₁₅X, as set forth in greater detail below.

in the active state by means other than binding of the receptor with its endogenous ligand or a chemical equivalent thereof. Preferably, a G protein-coupled receptor subjected to constitutive receptor activation in accordance with the invention disclosed herein evidences at least a 10% difference in response (increase or decrease, as the case may be) to the signal measured for

INSDOCID- WO 0022129A1 IA

15

constitutive activation as compared with the endogenous form of that GPCR, more preferably, about a 25% difference in such comparative response, and most preferably about a 50% difference in such comparative response. When used for the purposes of directly identifying candidate compounds, it is most preferred that the signal difference be at least about 50% such that there is a sufficient difference between the endogenous signal and the non-endogenous signal to differentiate between selected candidate compounds. In most instances, the "difference" will be an increase in signal; however, with respect to Gs-coupled GPCRS, the "difference" measured is preferably a decrease, as will be set forth in greater detail below.

CONTACT or CONTACTING shall mean bringing at least two moieties together, whether in an in vitro system or an in vivo system.

DIRECTLY IDENTIFYING or DIRECTLY IDENTIFIED, in relationship to the phrase "candidate compound", shall mean the screening of a candidate compound against a constitutively activated G protein-coupled receptor, and assessing the compound efficacy of such compound. This phrase is, under no circumstances, to be interpreted or understood to be encompassed by or to encompass the phrase "indirectly identifying" or "indirectly identified."

ENDOGENOUS shall mean a material that is naturally produced by the genome of the species. ENDOGENOUS in reference to, for example and not limitation, GPCR, shall mean that which is naturally produced by a human, an insect, a plant, a bacterium, or a virus. By contrast, the term NON-ENDOGENOUS in this context shall mean that which is not naturally produced by the genome of a species. For example, and not limitation, a receptor which is not constitutively active in its endogenous form, but when mutated by using the cassettes disclosed herein and thereafter becomes constitutively active, is most preferably referred to herein as a "non-endogenous, constitutively activated receptor." Both terms can be utilized to describe both "in

vivo" and "in vitro" systems. For example, and not limitation, in a screening approach, the endogenous or non-endogenous receptor may be in reference to an in vitro screening system whereby the receptor is expressed on the cell-surface of a mammalian cell. As a further example and not limitation, where the genome of a mammal has been manipulated to include a non-endogenous constitutively activated receptor, screening of a candidate compound by means of an in vivo system is viable.

therein. In the case of a prokaryotic Host Cell, a Plasmid is typically replicated as an autonomous molecule as the Host Cell replicates (generally, the Plasmid is thereafter isolated for introduction into a eukaryotic Host Cell); in the case of a eukaryotic Host Cell, a Plasmid is integrated into the cellular DNA of the Host Cell such that when the eukaryotic Host Cell replicates, the Plasmid replicates. Preferably, for the purposes of the invention disclosed herein, the Host Cell is eukaryotic, more preferably, mammalian, and most preferably selected from the group consisting of 293, 293T and COS-7 cells.

INDIRECTLY IDENTIFYING or INDIRECTLY IDENTIFIED means the traditional approach to the drug discovery process involving identification of an endogenous ligand specific for an endogenous receptor, screening of candidate compounds against the receptor for determination of those which interfere and/or compete with the ligand-receptor interaction, and assessing the efficacy of the compound for affecting at least one second messenger pathway associated with the activated receptor.

INHIBIT or INHIBITING, in relationship to the term "response" shall mean that a onse is decreased or prevented in the presence of a compound as opposed to in the absence of ound.

10

15

15

20

INVERSE AGONISTS shall mean compounds which bind to either the endogenous form of the receptor or to the constitutively activated form of the receptor, and which inhibit the baseline intracellular response initiated by the active form of the receptor below the normal base level of activity which is observed in the absence of agonists or partial agonists, or decrease GTP binding to membranes. Preferably, the baseline intracellular response is inhibited in the presence of the inverse agonist by at least 30%, more preferably by at least 50%, and most preferably by at least 75%, as compared with the baseline response in the absence of the inverse agonist.

KNOWN RECEPTOR shall mean an endogenous receptor for which the endogenous ligand specific for that receptor has been identified.

LIGAND shall mean an endogenous, naturally occurring molecule specific for an endogenous, naturally occurring receptor.

amino acid sequence shall mean a specified change or changes to such endogenous sequences such that a mutated form of an endogenous, non-constitutively activated receptor evidences constitutive activation of the receptor. In terms of equivalents to specific sequences, a subsequent mutated form of a human receptor is considered to be equivalent to a first mutation of the human receptor if (a) the level of constitutive activation of the subsequent mutated form of the receptor is substantially the same as that evidenced by the first mutation of the receptor; and (b) the percent sequence (amino acid and/or nucleic acid) homology between the subsequent mutated form of the receptor and the first mutation of the receptor is at least about 80%, more preferably at least about 90% and most preferably at least 95%. Ideally, and owing to the fact that the most preferred cassettes disclosed herein for achieving constitutive activation includes a single amino acid and/or codon change between the endogenous and the non-endogenous forms of the GPCR (i.e. X or

X_{codon}), the percent sequence homology should be at least 98%.

ORPHAN RECEPTOR shall mean an endogenous receptor for which the endogenous ligand specific for that receptor has not been identified or is not known.

PHARMACEUTICAL COMPOSITION shall mean a composition comprising at least one active ingredient, whereby the composition is amenable to investigation for a specified, efficacious outcome in a mammal (for example, and not limitation, a human). Those of ordinary skill in the art will understand and appreciate the techniques appropriate for determining whether an active ingredient has a desired efficacious outcome based upon the needs of the artisan.

PLASMID shall mean the combination of a Vector and cDNA. Generally, a Plasmid is
introduced into a Host Cell for the purpose of replication and/or expression of the cDNA as a protein.

STIMULATE or STIMULATING, in relationship to the term "response" shall mean that a response is increased in the presence of a compound as opposed to in the absence of the compound.

15 TRANSVERSE or TRANSVERSING, in reference to either a defined nucleic acid sequence or a defined amino acid sequence, shall mean that the sequence is located within at least two different and defined regions. For example, in an amino acid sequence that is 10 amino acid moieties in length, where 3 of the 10 moieties are in the TM6 region of a GPCR and the remaining 7 moieties are in the IC3 region of the GPCR, the 10 amino acid moiety can be described as transversing the TM6 and IC3 regions of the GPCR.

VECTOR in reference to cDNA shall mean a circular DNA capable of incorporating at least one cDNA and capable of incorporation into a Host Cell.

The order of the following sections is set forth for presentational efficiency and is not

WO 00/22129 PCT/US99/23938

- 19 -

intended, nor should be construed, as a limitation on the disclosure or the claims to follow.

A. Introduction

10

15

The traditional study of receptors has always proceeded from the a priori assumption (historically based) that the endogenous ligand must first be identified before discovery could proceed to find antagonists and other molecules that could affect the receptor. Even in cases where an antagonist might have been known first, the search immediately extended to looking for the endogenous ligand. This mode of thinking has persisted in receptor research even after the discovery of constitutively activated receptors. What has not been heretofore recognized is that it is the active state of the receptor that is most useful for discovering agonists, partial agonists, and inverse agonists of the receptor. For those diseases which result from an overly active receptor or an under-active receptor, what is desired in a therapeutic drug is a compound which acts to diminish the active state of a receptor or enhance the activity of the receptor, respectively, not necessarily a drug which is an antagonist to the endogenous ligand. This is because a compound that reduces or enhances the activity of the active receptor state need not bind at the same site as the endogenous ligand. Thus, as taught by a method of this invention, any search for therapeutic compounds should start by screening compounds against the ligand-independent active state.

Screening candidate compounds against non-endogenous, constitutively activated GPCRs allows for the direct identification of candidate compounds which act at these cell surface receptors, without requiring any prior knowledge or use of the receptor's endogenous ligand. By determining areas within the body where the endogenous version of such GPCRs are expressed and/or over-expressed, it is possible to determine related disease/disorder states which are associated with the expression and/or over-expression of these receptors; such an approach is disclosed in this patent document.

B. Disease/Disorder Identification and/r Selection

Most preferably, inverse agonists to the non-endogenous, constitutively activated GPCRs can be identified using the materials of this invention. Such inverse agonists are ideal candidates as lead compounds in drug discovery programs for treating diseases related to these receptors.

5 Because of the ability to directly identify inverse agonists, partial agonists or agonists to these receptors, thereby allowing for the development of pharmaceutical compositions, a search, for diseases and disorders associated with these receptors is possible. For example, scanning both diseased and normal tissue samples for the presence of these receptor now becomes more than an academic exercise or one which might be pursued along the path of identifying, in the case of an orphan receptor, an endogenous ligand. Tissue scans can be conducted across a broad range of healthy and diseased tissues. Such tissue scans provide a preferred first step in associating a specific receptor with a disease and/or disorder.

Preferably, the DNA sequence of the endogenous GPCR is used to make a probe for either radiolabeled cDNA or RT-PCR identification of the expression of the GPCR in tissue samples.

The presence of a receptor in a diseased tissue, or the presence of the receptor at elevated or decreased concentrations in diseased tissue compared to a normal tissue, can be preferably utilized to identify a correlation with that disease. Receptors can equally well be localized to regions of organs by this technique. Based on the known functions of the specific tissues to which the receptor is localized, the putative functional role of the receptor can be deduced.

20 C. A "Human GPCR Proline Marker" Algorithm and the Creation of Non-Endogenous, Constitutively-Active Human GPCRs

Among the many challenges facing the biotechnology arts is the unpredictability in gleaning genetic information from one species and correlating that information to another species

20

- nowhere in this art does this problem evidence more annoying exacerbation than in the genetic sequences that encode nucleic acids and proteins. Thus, for consistency and because of the highly unpredictable nature of this art, the following invention is limited, in terms of mammals, to human GPCRs – applicability of this invention to other mammalian species, while a potential possibility, is considered beyond mere rote application.

In general, when attempting to apply common "rules" from one related protein sequence to another or from one species to another, the art has typically resorted to sequence alignment, i.e., sequences are linearized and attempts are then made to find regions of commonality between two or more sequences. While useful, this approach does not always prove to result in meaningful to information. In the case of GPCRs, while the general structural motif is identical for all GPCRs, while the variations in lengths of the TMs, ECs and ICs make such alignment approaches from one GPCR to another difficult at best. Thus, while it may be desirable to apply a consistent approach to, e.g., constitutive activation from one GPCR to another, because of the great diversity in sequence length, fidelity, etc from one GPCR to the next, a generally applicable, and readily successful mutational alignment approach is in essence not possible. In an analogy, such an approach is akin to having a traveler start a journey at point A by giving the traveler dozens of different maps to point B, without any scale or distance markers on any of the maps, and then asking the traveler to find the shortest and most efficient route to destination B only by using the maps. In such a situation, the task can be readily simplified by having (a) a common "placemarker" on each map, and (b) the ability to measure the distance from the place-marker to destination B - this, then, will allow the traveler to select the most efficient from starting-point A to destination B.

In essence, a feature of the invention is to provide such coordinates within human GPCRs

that readily allows for creation of a constitutively active form of the human GPCRs.

As those in the art appreciate, the transmembrane region of a cell is highly hydrophobic; thus, using standard hydrophobicity plotting techniques, those in the art are readily able to determine the TM regions of a GPCR, and specifically TM6 (this same approach is also applicable to determining the EC and IC regions of the GPCR). It has been discovered that within the TM6 region of human GPCRs, a common proline residue (generally near the middle of TM6), acts as a constitutive activation "marker." By counting 15 amino acids from the proline marker, the 16th amino acid (which is located in the IC3 loop), when mutated from its endogenous form to a non-endogenous form, leads to constitutive activation of the receptor. For convenience, we refer to this as the "Human GPCR Proline Marker" Algorithm. Although the non-endogenous amino acid at this position can be any of the amino acids, most preferably, the non-endogenous amino acid is lysine. While not wishing to be bound by any theory, we believe that this position itself is unique and that the mutation at this location impacts the receptor to allow for constitutive activation.

We note that, for example, when the endogenous amino acid at the 16th position is already lysine (as is the case with GPR4 and GPR32), then in order for X to be a non-endogenous amino acid, it must be other than lysine; thus, in those situations where the endogenous GPCR has an endogenous lysine residue at the 16th position, the non-endogenous version of that GPCR preferably incorporates an amino acid other than lysine, preferably alanine, histidine and arginine, at this position. Of further note, it has been determined that GPR4 appears to be linked to Gs and active in its endogenous form (data not shown).

Because there are only 20 naturally occurring amino acids (although the use of non-naturally occurring amino acids is also viable), selection of a particular non-endogenous amino

10

15

20

acid for substitution at this 16th position is viable and allows for efficient selection of a non-endogenous amino acid that fits the needs of the investigator. However, as noted, the more preferred non-endogenous amino acids at the 16th position are lysine, hisitidine, arginine and alanine, with lysine being most preferred. Those of ordinary skill in the art are credited with the ability to readily determine proficient methods for changing the sequence of a codon to achieve a desired mutation.

It has also been discovered that occasionally, but not always, the proline residue marker will be preceded in TM6 by W2 (*i.e.*, W2P¹AA₁₅X) where W is tryptophan and 2 is any amino acid residue.

Our discovery, amongst other things, negates the need for unpredictable and complicated sequence alignment approaches commonly used by the art. Indeed, the strength of our discovery, while an algorithm in nature, is that it can be applied in a facile manner to human GPCRs, with dexterous simplicity by those in the art, to achieve a unique and highly useful end-product, *i.e.*, a constitutively activated version of a human GPCR. Because many years and significant amounts of money will be required to determine the endogenous ligands for the human GPCRs that the Human Genome project is uncovering, the disclosed invention not only reduces the time necessary to positively exploit this sequence information, but at significant cost-savings. This approach truly validates the importance of the Human Genome Project because it allows for the utilization of genetic information to not only understand the role of the GPCRs in, *e.g.*, diseases, but also provides the opportunity to improve the human condition.

D. Screening of Candidate Compounds

1. Generic GPCR screening assay techniques

When a G protein receptor becomes constitutively active, it couples to a G protein (e.g.,

Gq, Gs, Gi, Go) and stimulates release and subsequent binding of GTP to the G protein. The G protein then acts as a GTPase and slowly hydrolyzes the GTP to GDP, whereby the receptor, under normal conditions, becomes deactivated. However, constitutively activated receptors, including the non-endogenous, human constitutively active GPCRs of the present invention, continue to exchange GDP for GTP. A non-hydrolyzable analog of GTP, [35S]GTPγS, can be used to monitor enhanced binding to G proteins present on membranes which express constitutively activated receptors. It is reported that [35S]GTPγS can be used to monitor G protein coupling to membranes in the absence and presence of ligand. An example of this monitoring, among other examples well-known and available to those in the art, was reported by Traynor and Nahorski in 1995. The preferred use of this assay system is for initial screening of candidate compounds because the system is generically applicable to all G protein-coupled receptors regardless of the particular G protein that interacts with the intracellular domain of the receptor.

B 2. Specific GPCR screening assay techniques

C Once candidate compounds are identified using the "generic" G protein-coupled receptor assay (i.e., an assay to select compounds that are agonists, partial agonists, or inverse agonists), further screening to confirm that the compounds have interacted at the receptor site is preferred. For example, a compound identified by the "generic" assay may not bind to the receptor, but may instead merely "uncouple" the G protein from the intracellular domain.

a. Gs and Gi.

Gs stimulates the enzyme adenylyl cyclase. Gi (and Go), on the other hand, inhibit this enzyme. Adenylyl cyclase catalyzes the conversion of ATP to cAMP; thus,

15

WO 00/22129 PCT/US99/23938

constitutively activated GPCRs that couple the Gs protein are associated with increased cellular levels of cAMP. On the other hand, constitutively activated GPCRs that couple the Gi (or Go) protein are associated with decreased cellular levels of cAMP. See, generally, "Indirect Mechanisms of Synaptic Transmission," Chpt. 8, From Neuron To Brain (3rd Ed.) Nichols, J.G. et al eds. Sinauer Associates, Inc. (1992). Thus, assays that detect cAMP can be utilized to determine if a candidate compound is, e.g., an inverse agonist to the receptor (i.e., such a compound would decrease the levels of cAMP). A variety of approaches known in the art for measuring cAMP can be utilized; a most preferred approach relies upon the use of anti-cAMP antibodies in an ELISA-based format. Another type of assay that can be 10 utilized is a whole cell second messenger reporter system assay. Promoters on genes drive the expression of the proteins that a particular gene encodes. Cyclic AMP drives gene expression by promoting the binding of a cAMP-responsive DNA binding protein or transcription factor (CREB) which then binds to the promoter at specific sites called cAMP response elements and drives the expression of the gene. Reporter systems can be constructed which have a promoter containing multiple cAMP response elements before the reporter gene, e.g., β-galactosidase or luciferase. 15 Thus, a constitutively activated Gs-linked receptor causes the accumulation of cAMP that then activates the gene and expression of the reporter protein. The reporter protein such as βgalactosidase or luciferase can then be detected using standard biochemical assays (Chen et al. 1995). With respect to GPCRs that link to Gi (or Go), and thus decrease levels of cAMP, an 20 approach to the screening of, e.g., inverse agonists, based upon utilization of receptors that link to Gs (and thus increase levels of cAMP) is disclosed in the Example section with respect to GPR17 and GPR30.

than the state of the product of a grade of the population

b. Go and Gg.

Gq and Go are associated with activation of the enzyme phospholipase C, which in turn hydrolyzes the phospholipid PIP2, releasing two intracellular messengers: diacycloglycerol (DAG) and inistol 1,4,5-triphoisphate (IP₃). Increased accumulation of IP₃ is associated with activation of Gq- and Go-associated receptors. See, generally, "Indirect Mechanisms of Synaptic Transmission," Chpt. 8, From Neuron To Brain (3rd Ed.) Nichols, J.G. et al eds. Sinauer Associates, Inc. (1992). Assays that detect IP₃ accumulation can be utilized to determine if a candidate compound is, e.g., an inverse agonist to a Gq- or Goassociated receptor (i.e., such a compound would decrease the levels of IP₃). Gq-associated receptors can also been examined using an AP1 reporter assay in that Gq-dependent 10 phospholipase C causes activation of genes containing AP1 elements; thus, activated Gqassociated receptors will evidence an increase in the expression of such genes, whereby inverse agonists thereto will evidence a decrease in such expression, and agonists will evidence an increase in such expression. Commercially available assays for such detection are available.

Medicinal Chemistry

Generally, but not always, direct identification of candidate compounds is preferably conducted in conjunction with compounds generated via combinatorial chemistry techniques, whereby thousands of compounds are randomly prepared for such analysis. Generally, the results of such screening will be compounds having unique core structures; thereafter, these compounds are preferably subjected to additional chemical modification around a preferred core structure(s) to further enhance the medicinal properties thereof. Such techniques are

15

•WO 00/22129 PCT/US99/23938

- 27 -

known to those in the art and will not be addressed in detail in this patent document.

F. Pharmaceutical Compositions

Candidate compounds selected for further development can be formulated into pharmaceutical compositions using techniques well known to those in the art. Suitable pharmaceutically-acceptable carriers are available to those in the art; for example, *see* Remington's Pharmaceutical Sciences, 16th Edition, 1980, Mack Publishing Co., (Oslo et al., eds.)

G. Other Utility

Although a preferred use of the non-endogenous versions of the disclosed human GPCRs is for the direct identification of candidate compounds as inverse agonists, agonists or partial agonists (preferably for use as pharmaceutical agents), these receptors can also be utilized in research settings. For example, in vitro and in vivo systems incorporating these receptors can be utilized to further elucidate and understand the roles of the receptors in the human condition, both normal and diseased, as well understanding the role of constitutive activation as it applies to understanding the signaling cascade. A value in these non-endogenous receptors is that their utility as a research tool is enhanced in that, because of their unique features, the disclosed receptors can be used to understand the role of a particular receptor in the human body before the endogenous ligand therefor is identified. Other uses of the disclosed receptors will become apparent to those in the art based upon, *inter alia*, a review of this patent document.

EXAMPLES

The following examples are presented for purposes of elucidation, and not limitation, of the present invention. Following the teaching of this patent document that a mutational cassette may be utilized in the IC3 loop of human GPCRs based upon a position relative to a proline residue in TM6 to constitutively activate the receptor, and while specific nucleic acid

and amino acid sequences are disclosed herein, those of ordinary skill in the art are credited with the ability to make minor modifications to these sequences while achieving the same or substantially similar results reported below. Particular approaches to sequence mutations are within the purview of the artisan based upon the particular needs of the artisan.

5 Example 1 Preparation of Endogenous Human GPCRs

A variety of GPCRs were utilized in the Examples to follow. Some endogenous human GPCRs were graciously provided in expression vectors (as acknowledged below) and other endogenous human GPCRs were synthesized *de novo* using publicly-available sequence information.

1. GPR1 (GenBank Accession Number: U13666)

The human cDNA sequence for GPR1 was provided in pRcCMV by Brian O'Dowd (University of Toronto). GPR1 cDNA (1.4kB fragment) was excised from the pRcCMV vector as a NdeI-Xbal fragment and was subcloned into the NdeI-Xbal site of pCMV vector (see Figure 3). Nucleic acid (SEQ.ID.NO.: 1) and amino acid (SEQ.ID.NO.: 2) sequences for human GPR1 were thereafter determined and verified.

2. GPR4 (GenBank Accession Numbers: L36148, U35399, U21051)

The human cDNA sequence for GPR4 was provided in pRcCMV by Brian O'Dowd (University of Toronto). GPR1 cDNA (1.4kB fragment) was excised from the pRcCMV vector as an ApaI(blunted)-XbaI fragment and was subcloned (with most of the 5' untranslated region removed) into HindIII(blunted)-XbaI site of pCMV vector. Nucleic acid (SEQ.ID.NO.: 3) and amino acid (SEQ.ID.NO.: 4) sequences for human GPR4 were thereafter determined and verified.

3. GPR5 (GenBank Accession Number: L36149)

BNSDCCID: <WO 0022129A1 IA>

20

The cDNA for human GPR5 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 64°C for 1 min; and 72°C for 1.5 min. The 5' PCR primer contained an EcoRI site with the sequence: 5'-TATGAATTCAGATGCTCTAAACGTCCCTGC-3' (SEQ.ID.NO.: 5) and the 3' primer contained BamHI site with the sequence: 5'-TCCGGATCCACCTGCACCTGCGCCTGCACC-3' (SEQ.ID.NO.: 6). The 1.1 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of PCMV expression vector. Nucleic acid (SEQ.ID.NO.: 7) and amino acid (SEQ.ID.NO.: 8) sequences for human GPR5 were thereafter determined and verified.

4. GPR7 (GenBank Accession Number: U22491)

The cDNA for human GPR7 was generated and cloned into pCMV expression vector as follows: PCR condition- PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1 min; and 72°C for 1 min and 20 sec. The 5' PCR primer contained a HindIII site with the sequence:

- 5'-GCAAGCTTGGGGGACGCCAGGTCGCCGGCT-3' (SEQ.ID.NO.: 9)
- and the 3' primer contained a BamHI site with the sequence:
 - 5'-GCGGATCCGGACGCTGGGGGAGTCAGGCTGC-3' (SEQ.ID.NO.: 10).
 - The 1.1 kb PCR fragment was digested with HindIII and BamHI and cloned into HindIII-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 11) and amino acid (SEQ.ID.NO.:

12) sequences for human GPR7 were thereafter determined and verified.

5. GPR8 (GenBank Accession Number: U22492)

The cDNA for human GPR8 was generated and cloned into pCMV expression

vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1 min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the sequence:

- -5'-CGGAATTCGTCAACGGTCCCAGCTACAATG-3' (SEQ.ID.NO.: 13).
- and the 3' primer contained a BamHI site with the sequence:
 - 5'-ATGGATCCCAGGCCCTTCAGCACCGCAATAT-3'(SEQ.ID.NO.: 14).

The 1.1 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of PCMV expression vector. All 4 cDNA clones sequenced contained a possible polymorphism involving a change of amino acid 206 from Arg to Gln. Aside from this difference, nucleic acid (SEQ.ID.NO.: 15) and amino acid (SEQ.ID.NO.: 16) sequences for human GPR8 were thereafter determined and verified.

6. GPR9 (GenBank Accession Number: X95876)

The cDNA for human GPR9 was generated and cloned into pCMV expression vector as follows: PCR was performed using a clone (provided by Brian O'Dowd) as template and pfu polymerase (Stratagene) with the buffer system provided by the manufacturer supplemented with 10% DMSO, 0.25 µM of each primer, and 0.5 mM of each of the 4 nucleotides. The cycle condition was 25 cycles of: 94°C for 1 min; 56°C for 1min; and 72 °C for 2.5 min. The 5' PCR primer contained an EcoRI site with the sequence:

PCT/US99/23938

20

5'-ACGAATTCAGCCATGGTCCTTGAGGTGAGTGACCACCAAGTGCTAAAT-3' (SEQ.ID.NO.: 17)

and the 3' primer contained a BamHI site with the sequence:

5'-GAGGATCCTGGAATGCGGGGAAGTCAG-3' (SEQ.ID.NO.: 18).

The 1.2 kb PCR fragment was digested with EcoRI and cloned into EcoRI-Smal site of PCMV expression vector. Nucleic acid (SEQ.ID.NO.: 19) and amino acid (SEQ.ID.NO.: 20) sequences for human GPR9 were thereafter determined and verified.

7. GPR9-6 (GenBank Accession Number: U45982)

The cDNA for human GPR9-6 was generated and cloned into pCMV expression.

10 vector as follows: PCR was performed using genomic DNA as template and rTth polymerases:

(Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1 min; and 72 °C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence: 5'-TTAAGCTTGACCTAATGCCATCTTGTGTCC-3' (SEQ.ID.NO.: 21)

and the 3' primer contained a BamHI site with the sequence:

5'-TTGGATCCAAAAGAACCATGCACCTCAGAG-3' (SEQ.ID.NO.: 22).

The 1.2 kb PCR fragment was digested with BamHI and cloned into EcoRV-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 23) and amino acid (SEQ.ID.NO.: 24) sequences for human GPR9-6 were thereafter determined and verified.

8. GPR10 (GenBank Accession Number: U32672)

The human cDNA sequence for GPR10 was provided in pRcCMV by Brian O'Dowd (University of Toronto). GPR10 cDNA (1.3kB fragment) was excised from the pRcCMV vector as an EcoRI-XbaI fragment and was subcloned into EcoRI-XbaI site of pCMV

vector. Nucleic acid (SEQ.ID.NO.: 25) and amino acid (SEQ.ID.NO.: 26) sequences for human GPR10 were thereafter determined and verified.

9. GPR15 (GenBank Accession Number: U34806)

The human cDNA sequence for GPR15 was provided in pCDNA3 by Brian O'Dowd (University of Toronto). GPR15 cDNA (1.5kB fragment) was excised from the pCDNA3 vector as a HindIII-Bam fragment and was subcloned into HindIII-Bam site of pCMV vector. Nucleic acid (SEQ.ID.NO.: 27) and amino acid (SEQ.ID.NO.: 28) sequences for human GPR15 were thereafter determined and verified.

10. GPR17 (GenBank Accession Number: Z94154)

- The cDNA for human GPR17 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 56°C for 1 min and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the sequence:
 - 5'-CTAGAATTCTGACTCCAGCCAAAGCATGAAT-3' (SEQ.ID.NO.: 29) and the 3' primer contained a BamHI site with the sequence:
 - 5'-GCTGGATCCTAAACAGTCTGCGCTCGGCCT-3' (SEQ.ID.NO.: 30).
- The 1.1 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI
 site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 31) and amino acid (SEQ.ID.NO.: 32) sequences for human GPR17 were thereafter determined and verified.

11. GPR18 (GenBank Accession Number: L42324)

The cDNA for human GPR18 was generated and cloned into pCMV expression

שאבטטטיט -MU טואטואפאי וא-

vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 54°C for 1 min; and 72°C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence:

5'-ATAAGATGATCACCCTGAACAATCAAGAT -3' (SEQ.ID.NO.: 33)
and the 3' primer contained an EcoRI site with the sequence:

5'-TCCGAATTCATAACATTTCACTGTTTATATTGC-3' (SEQ.ID.NO.: 34).

The 1.0 kb PCR fragment was digested with EcoRI and cloned into blunt-EcoRI site of pCMV expression vector. All 8 cDNA clones sequenced contained 4 possible polymorphisms involving changes of amino acid 12 from Thr to Pro, amino acid 86 from Ala to Glu, amino acid 97 from Ile to Leu and amino acid 310 from Leu to Met. Aside from these changes, nucleic acid (SEQ.ID.NO.: 35) and amino acid (SEQ.ID.NO.: 36) sequences for human GPR18 were thereafter determined and verified.

12. GPR20 (GenBank Accession Number: U66579)

The cDNA for human GPR20 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1 min; and 72 °C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence:

5'-CCAAGCTTCCAGGCCTGGGGTGTGCTGG-3' (SEQ.ID.NO.: 37)
and the 3' primer contained a BamHI site with the sequence:
5'-ATGGATCCTGACCTTCGGCCCCTGGCAGA-3' (SEQ.ID.NO.: 38).

The 1.2 kb PCR fragment was digested with BamHI and cloned into EcoRV-BamHI site of

PCMV expression vector. Nucleic acid (SEQ.ID.NO.: 39) and amino acid (SEQ.ID.NO.: 40) sequences for human GPR20 were thereafter determined and verified.

13. GPR21 (GenBank Accession Number: U66580)

The cDNA for human GPR21 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1 min; and 72 °C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence: 5'-GAGAATTCACTCCTGAGCTCAAGATGAACT-3' (SEQ.ID.NO.: 41)

and the 3' primer contained a BamHI site with the sequence:

5'-CGGGATCCCCGTAACTGAGCCACTTCAGAT-3' (SEQ.ID.NO.: 42).

The 1.1 kb PCR fragment was digested with BamHI and cloned into EcoRV-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 43) and amino acid (SEQ.ID.NO.: 44) sequences for human GPR21 were thereafter determined and verified.

14. GPR22 (GenBank Accession Number: U66581)

The cDNA for human GPR22 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 50°C for 1 min; 50°C

20 for 1min; and 72 °C for 1.5 min. The 5' PCR primer was kinased with the sequence:

5'-TCCCCGGGAAAAAAACCAACTGCTCCAAA-3' (SEQ.ID.NO.: 45)

and the 3' primer contained a BamHI site with the sequence:

5'-TAGGATCCATTTGAATGTGGATTTGGTGAAA-3' (SEQ.ID.NO.: 46).

15

The 1.38 kb PCR fragment was digested with BamHI and cloned into EcoRV-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 47) and amino acid (SEQ.ID.NO.: 48) sequences for human GPR22 were thereafter determined and verified.

15. GPR24 (GenBank Accession Number: U71092)

The cDNA for human GPR24 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 56°C for 1 min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contains a HindIII site with the sequence:

5'-GTGAAGCTTGCCTCTGGTGCCTGCAGGAGG-3' (SEQ.ID.NO.: 49) and the 3' primer contains an EcoRI site with the sequence:
5'-GCAGAATTCCCGGTGGCGTGTTGTGGTGCCC-3' (SEQ.ID.NO.: 50).

The 1.3 kb PCR fragment was digested with HindIII and EcoRI and cloned into HindIII-EcoRI site of pCMV expression vector. The nucleic acid (SEQ.ID.NO.: 51) and amino acid sequence (SEQ.ID.NO.: 52) for human GPR24 were thereafter determined and verified.

16. GPR30 (GenBank Accession Number: U63917)

The cDNA for human GPR30 was generated and cloned as follows: the coding sequence of GPR30 (1128bp in length) was amplified from genomic DNA using the primers:

- 5'-GGCGGATCCATGGATGTGACTTCCCAA-3' (SEQ.ID.NO.: 53) and 5'-GGCGGATCCCTACACGGCACTGCTGAA-3' (SEQ.ID.NO.: 54).
 - The amplified product was then cloned into a commercially available vector, pCR2.1 (Invitrogen),

using a "TOPO-TA Cloning Kit" (Invitrogen, #K4500-01), following manufacturer instructions. The full-length GPR30 insert was liberated by digestion with BamH1, separated from the vector by agarose gel electrophoresis, and purified using a Sephaglas Bandprep™ Kit (Pharmacia, #27-9285-01) following manufacturer instructions. The nucleic acid (SEQ.ID.NO.: 55) and amino acid sequence (SEQ.ID.NO.: 56) for human GPR30 were thereafter determined and verified.

17. GPR31 (GenBank Accession Number: U65402)

The cDNA for human GPR31 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 58°C for 1 min; and 72 °C for 2 min. The 5' PCR primer contained an EcoRI site with the sequence: 5'-AAGGAATTCACGGCCGGGTGATGCCATTCCC-3' (SEQ.ID.NO.: 57) and the 3' primer contained a BamHI site with the sequence: 5'-GGTGGATCCATAAACACGGGCGTTGAGGAC -3' (SEQ.ID.NO.: 58).

The 1.0 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 59) and amino acid (SEQ.ID.NO.: 60) sequences for human GPR31 were thereafter determined and verified.

18. GPR32 (GenBank Accession Number: AF045764)

The cDNA for human GPR32 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 56°C for 1 min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the

garage and a transfer of the contract was a second

sequence:

5'-TAAGAATTCCATAAAAATTATGGAATGG-3' (SEQ.ID.NO.:243)
and the 3' primer contained a BamHI site with the sequence:
5'-CCAGGATCCAGCTGAAGTCTTCCATCATTC-3' (SEQ.ID.NO.: 244).

The 1.1 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 245) and amino acid (SEQ.ID.NO.: 246) sequences for human GPR32 were thereafter determined and verified.

19. GPR40 (GenBank Accession Number: AF024687)

The cDNA for human GPR40 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min, 65°C for 1 min and 72 °C for 1 min and 10 sec. The 5' PCR primer contained an EcoRI site with the sequence

5'-GCAGAATTCGGCGGCCCCATGGACCTGCCCCC-3' (SEQ.ID.NO.: 247)
and the 3' primer contained a BamHI site with the sequence
5'-GCTGGATCCCCCGAGCAGTGGCGTTACTTC-3' (SEQ.ID.NO.: 248).
The 1 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site
of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 249) and amino acid (SEQ.ID.NO.: 250)
sequences for human GPR40 were thereafter determined and verified.

20. GPR41 (GenBank Accession Number AF024688)

The cDNA for human GPR41 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase

(Perkin Elmer) with the buffer system provided by the manufacturer, $0.25~\mu\text{M}$ of each primer, and 0.2~mM of each 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 65°C for 1 min and 72 °C for 1 min and 10 sec. The 5' PCR primer contained an HindIII site with the sequence:

- 5 5'-CTCAAGCTTACTCTCTCACCAGTGGCCAC-3' (SEQ.ID.NO.: 251) and the 3' primer was kinased with the sequence 5'-CCCTCCTCCCCGGAGGACCTAGC-3' (SEQ.ID.NO.: 252).
 - The 1 kb PCR fragment was digested with HindIII and cloned into HindIII-blunt site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 253) and amino acid (SEQ.ID.NO.: 254) sequences for human GPR41 were thereafter determined and verified.

21. GPR43 (GenBank Accession Number AF024690)

The cDNA for human GPR43 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1 min; and 72 °C for 1 min and 10 sec. The 5' PCR primer contains an HindIII site with the sequence:

- 5'-TTTAAGCTTCCCCTCCAGGATGCTGCCGGAC-3' (SEQ.ID.NO.: 255) and the 3' primer contained an EcoRI site with the sequence:
- 5'-GGCGAATTCTGAAGGTCCAGGGAAACTGCTA-3' (SEQ.ID.NO. 256).
 The 1 kb PCR fragment was digested with HindIII and EcoRI and cloned into HindIII-EcoRI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 257) and amino acid (SEQ.ID.NO.: 258) sequences for human GPR43 were thereafter determined and verified.

22. APJ (GenBank Accession Number: U03642)

Human APJ cDNA (in pRcCMV vector) was provided by Brian O'Dowd (University of Toronto). The human APJ cDNA was excised from the pRcCMV vector as an EcoRI-Xbal (blunted) fragment and was subcloned into EcoRI-Smal site of pCMV vector. Nucleic acid (SEQ.ID.NO.: 61) and amino acid (SEQ.ID.NO.: 62) sequences for human APJ were thereafter determined and verified.

23. BLR1 (GenBank Accession Number: X68149)

The cDNA for human BLR1 was generated and cloned into pCMV expression vector as follows: PCR was performed using thymus cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1 min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the sequence:

- 5'-TGAGAATTCTGGTGACTCACAGCCGGCACAG-3' (SEQ.ID.NO.: 63):
- and the 3' primer contained a BamHI site with the sequence:
 - 5'-GCCGGATCCAAGGAAAAGCAGCAATAAAAGG-3' (SEQ.ID.NO.: 64). The 1.2 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 65) and amino acid (SEQ.ID.NO.: 66) sequences for human BLR1 were thereafter determined and verified.

24. CEPR (GenBank Accession Number: U77827)

The cDNA for human CEPR was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and

0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1 min; and 72 °C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence: 5'-CAAAGCTTGAAAGCTGCACGGTGCAGAGAC-3' (SEQ.ID.NO.:67) and the 3' primer contained a BamHI site with the sequence:

5'-GCGGATCCCGAGTCACACCCTGGCTGGGCC-3' (SEQ.ID.NO.: 68).

The 1.2 kb PCR fragment was digested with BamHI and cloned into EcoRV-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 69) and amino acid (SEQ.ID.NO.: 70) sequences for human CEPR were thereafter determined and verified.

25. EBI1 (GenBank Accession Number: L31581)

The cDNA for human EBI1 was generated and cloned into pCMV expression vector as follows: PCR was performed using thymus cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1 min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the sequence:

5'-ACAGAATTCCTGTGTGTGTTTTACCGCCCAG-3' (SEQ.ID.NO.: 71) and the 3' primer contained a BamHI site with the sequence:

5'-CTCGGATCCAGGCAGAAGAGTCGCCTATGG-3' (SEQ.ID.NO.: 72).

The 1.2 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of PCMV expression vector. Nucleic acid (SEQ.ID.NO.: 73) and amino acid (SEQ.ID.NO.: 74) sequences for human EBI1 were thereafter determined and verified.

26. EBI2 (GenBank Accession Number: L08177)

The cDNA for human EBI2 was generated and cloned into pCMV expression

vector as follows: PCR was performed using cDNA clone (graciously provided by Kevin Lynch, University of Virginia Health Sciences Center; the vector utilized was not identified by the source) as template and pfu polymerase (Stratagene) with the buffer system provided by the manufacturer supplemented with 10% DMSO, 0.25 μM of each primer, and 0.5 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 60°C for 1 min; and 72 °C for 1 min and 20 scc. The 5' PCR primer contained an EcoRI site with the sequence: 5'-CTGGAATTCACCTGGACCACCACCAATGGATA-3' (SEQ.ID.NO.: 75) and the 3' primer contained a BamHI site with the sequence 5'-CTCGGATCCTGCAAAGTTTGTCATACAG TT-3' (SEQ.ID.NO.: 76).

The 1.2 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 77) and amino acid (SEQ.ID.NO.: 78) sequences for human EBI2 were thereafter determined and verified.

27. ETBR-LP2 (GenBank Accession Number: D38449)

The cDNA for human ETBR-LP2 was generated and cloned into pCMV expression vector as follows: PCR was performed using brain cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1 min; and 72 °C for 1.5 min. The 5' PCR contained an EcoRI site with the sequence:

5'-CTGGAATTCTCCTGCTCATCCAGCCATGCGG -3' (SEQ.ID.NO.: 79)
and the 3' primer contained a BamHI site with the sequence:
5'-CCTGGATCCCCACCCCTACTGGGGCCTCAG -3' (SEQ.ID.NO.: 80).
The 1.5 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI

site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 81) and amino acid (SEQ.ID.NO.: 82) sequences for human ETBR-LP2 were thereafter determined and verified.

28. GHSR (GenBank Accession Number: U60179)

The cDNA for human GHSR was generated and cloned into pCMV expression vector as follows: PCR was performed using hippocampus cDNA as template and TaqPlus Precision polymerase (Stratagene) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 68°C for 1 min; and 72 °C for 1 min and 10 sec. For first round PCR, the 5' PCR primer sequence was:

- 10 5'-ATGTGGAACGCGACGCCCAGCG-3' (SEQ.ID.NO.: 83)
 - and the 3' primer sequence was:
 - 5'-TCATGTATTAATACTAGATTCT-3' (SEQ.ID.NO.: 84).
 - Two microliters of the first round PCR was used as template for the second round PCR where the 5' primer was kinased with sequence:

- 15 5'-TACCATGTGGAACGCGACGCCCAGCGAAGAGCCGGGGT-3'(SEQ.ID.NO.:85) and the 3' primer contained an EcoRI site with the sequence:
 - 5'-CGGAATTCATGTATTAATACTAGATTCTGTCCAGGCCCG-3'(SEQ.ID.NO.:86).
 - The 1.1 kb PCR fragment was digested with EcoRI and cloned into blunt-EcoRI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 87) and amino acid (SEQ.ID.NO.: 88) sequences
- 20 for human GHSR were thereafter determined and verified.

29. GPCR-CNS (GenBank Accession Number: AFO17262)

The cDNA for human GPCR-CNS was generated and cloned into pCMV expression vector as follows: PCR was performed using brain cDNA as template and rTth

polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, $0.25 \,\mu\text{M}$ of each primer, and $0.2 \,\text{mM}$ of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1 min; and 72°C for 2 min. The 5' PCR primer contained a HindIII site with the sequence:

- 5 5'-GCAAGCTTGTGCCCTCACCAAGCCATGCGAGCC-3' (SEQ.ID.NO.: 89)
 and the 3' primer contained an EcoRI site with the sequence:
 5'-CGGAATTCAGCAATGAGTTCCGACAGAAGC-3' (SEQ.ID.NO.: 90).
- The 1.9 kb PCR fragment was digested with HindIII and EcoRI and cloned into HindIII-EcoRI site of pCMV expression vector. All nine clones sequenced contained a potential polymorphism involving a S284C change. Aside from this difference, nucleic acid (SEQ.ID.NO.: 91) and amino acid (SEQ.ID.NO.: 92) sequences for human GPCR-CNS were thereafter determined and verified.

30. GPR-NGA (GenBank Accession Number: U55312)

The cDNA for human GPR-NGA was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 56°C for 1 min and 72°C for 1.5 min. The 5' PCR primer contained an EcoRI site with the sequence:

- 5'-CAGAATTCAGAGAAAAAAAGTGAATATGGTTTTT-3' (SEQ.ID.NO.: 93)
- 20 and the 3' primer contained a BamHI site with the sequence:
 - 5'-TTGGATCCCTGGTGCATAACAATTGAAAGAAT-3' (SEQ.ID.NO.: 94).
 - The 1.3 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 95) and amino acid (SEQ.ID.NO.:

96) sequences for human GPR-NGA were thereafter determined and verified.

31. H9 (GenBank Accession Number: U52219)

The cDNA for human HB954 was generated and cloned into pCMV expression vector as follows: PCR was performed using pituitary cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min, 62°C for 1 min and 72°C for 2 min. The 5' PCR primer contains a HindIII site with the sequence: 5'-GGAAAGCTTAACGATCCCCAGGAGCAACAT-3' (SEQ.ID.NO.: 97) and the 3' primer contains a BamHI site with the sequence:

10 5'-CTGGGATCCTACGAGAGCATTTTTCACACAG-3' (SEQ.ID.NO.: 98).

The 1.9 kb PCR fragment was digested with HindlII and BamHI and cloned into HindIII-BamHI site of pCMV expression vector. When compared to the published sequences, a different isoform with 12 bp in frame insertion in the cytoplasmic tail was also identified and designated "H9b." Both isoforms contain two potential polymorphisms involving changes of amino acid P320S and amino acid G448A. Isoform H9a contained another potential polymorphism of amino acid S493N, while isoform H9b contained two additional potential polymorphisms involving changes of amino acid I502T and amino acid A532T (corresponding to amino acid 528 of isoform H9a). Nucleic acid (SEQ.ID.NO.: 99) and amino acid (SEQ.ID.NO.: 100) sequences for human H9 were thereafter determined and verified (in the section below, both isoforms were mutated in accordance with the Human GPCR Proline Marker Algorithm).

32. HB954 (GenBank Accession Number: D38449)

The cDNA for human HB954 was generated and cloned into pCMV expression

vector as follows: PCR was performed using brain cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 58°C for 1 min and 72 °C for 2 min. The 5' PCR contained a HindIII site with the sequence:

- 5'-TCCAAGCTTCGCCATGGGACATAACGGGAGCT -3' (SEQ:ID.NO.: 101)
 and the 3' primer contained an EcoRI site with the sequence:
 5'-CGTGAATTCCAAGAATTTACAATCCTTGCT -3' (SEQ:ID.NO.: 102).
 - The 1.6 kb PCR fragment was digested with HindIII and EcoRI and cloned into HindIII-EcoRI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 103) and amino acid (SEQ.ID.NO.: 104) sequences for human HB954 were thereafter determined and verified.

33. HG38 (GenBank Accession Number: AF062006)

The cDNA for human HG38 was generated and cloned into pCMV expression vector as follows: PCR was performed using brain cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 56°C for 1 min and 72°C for 1 min and 30 sec. Two PCR reactions were performed to separately obtain the 5' and 3' fragment. For the 5' fragment, the 5' PCR primer contained an HindIII site with the sequence: 5'-CCCAAGCTTCGGGCACCATGGACACCTCCC-3' (SEQ.ID.NO.: 259) and the 3' primer contained a BamHIsite with the sequence:

20 5'-ACAGGATCCAAATGCACAGCACTGGTAAGC-3' (SEQ.ID.NO.: 260).

This 5' 1.5 kb PCR fragment was digested with HindIII and BamHI and cloned into an HindIII-BamHI site of pCMV. For the 3' fragment, the 5' PCR primer was kinased with the sequence: 5'-CTATAACTGGGTTACATGGTTTAAC-3' (SEQ.ID.NO. 261)

and the 3' primer contained an EcoRI site with the sequence:

5'-TTTGAATTCACATATTAATTAGAGACATGG-3' (SEQ.ID.NO.: 262).

The 1.4 kb 3' PCR fragment was digested with EcoRI and subcloned into a blunt-EcoRI site of pCMV vector. The 5' and 3' fragments were then ligated together through a common EcoRV site to generate the full length cDNA clone. Nucleic acid (SEQ.ID.NO.: 263) and amino acid (SEQ.ID.NO.: 264) sequences for human HG38 were thereafter determined and verified.

34. HM74 (GenBank Accession Number: D10923)

The cDNA for human HM74 was generated and cloned into pCMV expression vector as follows: PCR was performed using either genomic DNA or thymus cDNA (pooled) as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1 min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the sequence:

- 5'-GGAGAATTCACTAGGCGAGGCGCTCCATC-3' (SEQ.ID.NO.: 105)
- 15 and the 3' primer was kinased with the sequence:
 - 5'-GGAGGATCCAGGAAACCTTAGGCCGAGTCC-3' (SEQ.ID.NO.:106).

The 1.3 kb PCR fragment was digested with EcoRI and cloned into EcoRI-SmaI site of pCMV expression vector. Clones sequenced revealed a potential polymorphism involving a N94K change. Aside from this difference, nucleic acid (SEQ.ID.NO.: 107) and amino acid (SEQ.ID.NO.: 108) sequences for human HM74 were thereafter determined and verified.

35. MIG (GenBank Accession Numbers: AFO44600 and AFO44601)

The cDNA for human MIG was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and TaqPlus Precision

RNSDOCID -WO noostoost is-

20

—III

polymerase (Stratagene) for first round PCR or pfu polymerase (Stratagene) for second round PCR with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM (TaqPlus Precision) or 0.5 mM (pfu) of each of the 4 nucleotides. When pfu was used, 10% DMSO was included in the buffer. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1 min; and 72 °C for: (a) 1 min for first round PCR; and (b) 2 min for second round PCR. Because there is an intron in the coding region, two sets of primers were separately used to generate overlapping 5' and 3' fragments. The 5' fragment PCR primers were:

5'-ACCATGGCTTGCAATGGCAGTGCGGCCAGGGGGCACT-3' (external sense) (SEQ.ID.NO.: 109)

- 10 and
 - 5'-CGACCAGGACAAACAGCATCTTGGTCACTTGTCTCCGGC-3'(internal antisense) (SEQ.ID.NO.: 110).
 - The 3' fragment PCR primers were:
 - 5'-GACCAAGATGCTGTTTGTCCTGGTCGTGGTGTTTTGGCAT-3' (internal sense)
- 15 (SEQ.ID.NO.: 111) and
 - 5'-CGGAATTCAGGATGGATCGGTCTCTTGCTGCGCCT-3' (external antisense with an EcoRI site) (SEQ.ID.NO.: 112).
- The 5' and 3' fragments were ligated together by using the first round PCR as template and the kinased external sense primer and external antisense primer to perform second round PCR. The 1.2 kb PCR fragment was digested with EcoRI and cloned into the blunt-EcoRI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 113) and amino acid (SEQ.ID.NO.: 114) sequences for human MIG were thereafter determined and verified.

36. OGR1 (GenBank Accession Number: U48405)

The cDNA for human OGR1 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1 min; and 72 °C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence: 5'-GGAAGCTTCAGGCCCAAAGATGGGGAACAT-3' (SEQ.ID.NO.: 115): and the 3' primer contained a BamHI site with the sequence: 5'-GTGGATCCACCCGGGGAGGACCCAGGCTAG -3' (SEQ.ID.NO.: 116).

The 1.1 kb PCR fragment was digested with BamHI and cloned into the EcoRV-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 117) and amino acid (SEQ.ID.NO.: 118) sequences for human OGR1 were thereafter determined and verified.

37. Serotonin 5HT_{2A}

The cDNA encoding endogenous human 5HT_{2A} receptor was obtained by RT-PCR using human brain poly-A⁺ RNA; a 5' primer from the 5' untranslated region with an Xho I restriction site:

5'-GACCTCGAGTCCTTCTACACCTCATC-3' (SEQ.ID.NO: 119) and a 3' primer from the 3' untranslated region containing an Xba I site: 5'-TGCTCTAGATTCCAGATAGGTGAAAACTTG-3' (SEQ.ID.NO: 120)

PCR was performed using either TaqPlusTM precision polymerase (Stratagene) or rTthTM polymerase (Perkin Elmer) with the buffer system provided by the manufacturers, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 57°C for 1 min; and 72°C for 2 min. The 1.5 kb PCR fragment was digested with Xba I and subcloned into Eco RV-Xba I site of pBluescript. The resulting cDNA clones were fully

BNSDOCID <WO DOSSISCATIA

15

sequenced and found to encode two amino acid changes from the published sequences. The first one was a T25N mutation in the N-terminal extracellular domain; the second is an H452Y mutation. Because cDNA clones derived from two independent PCR reactions using Taq polymerase from two different commercial sources (TaqPlusTM from Stratagene and rTthTM Perkin Elmer) contained the same two mutations, these mutations are likely to represent sequence polymorphisms rather than PCR errors. With these exceptions, the nucleic acid (SEQ.ID.NO.: 121) and amino acid (SEQ.ID.NO.: 122) sequences for human 5HT_{2A} were thereafter determined and verified.

38. Serotonin 5HT_{2C}

The cDNA encoding endogenous human 5HT_{2C} receptor was obtained from human brain poly-A⁺ RNA by RT-PCR. The 5' and 3' primers were derived from the 5' and 3' untranslated regions and contained the following sequences:

5'-GACCTCGAGGTTGCTTAAGACTGAAGC-3' (SEQ.ID.NO.: 123)

5'-ATTTCTAGACATATGTAGCTTGTACCG-3' (SEQ.ID.NO.: 124)

Nucleic acid (SEQ.ID.NO.: 125) and amino acid (SEQ.ID.NO.: 126) sequences for human 5HT_{2C} were thereafter determined and verified.

39. V28 (GenBank Accession Number: U20350)

The cDNA for human V28 was generated and cloned into pCMV expression vector as follows: PCR was performed using brain cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1 min; and 72°C for 1 min and 20 sec. The 5' PCR primer contained a HindIII site with the sequence: 5'-GGTAAGCTTGGCAGTCCACGCCAGGCCTTC-3' (SEQ.ID.NO.: 127)

and the 3' primer contained an EcoRI site with the sequence:

5'-TCCGAATTCTCTGTAGACACAAGGCTTTGG-3' (SEQ.ID.NO.: 128)

The 1.1 kb PCR fragment was digested with HindIII and EcoRI and cloned into HindIII-EcoRI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 129) and amino acid (SEQ.ID.NO.:

5 130) sequences for human V28 were thereafter determined and verified.

Example 2
PREPARATION OF NON-ENDOGENOUS HUMAN GPCRS

1. Site-Directed Mutagenesis

Mutagenesis based upon the Human GPCR Proline Marker approach disclosed herein was

10 performed on the foregoing endogenous human GPCRs using Transformer Site-Directed

Mutagenesis Kit (Clontech) according to the manufacturer instructions. For this mutagenesis

approach, a Mutation Probe and a Selection Marker Probe (unless otherwise indicated, the probe

of SEQ.ID.NO.: 132 was the same throughout) were utilized, and the sequences of these for the

specified sequences are listed below in Table B (the parenthetical number is the SEQ. ID.NO.).

15 For convenience, the codon mutation incorporated into the human GPCR is also noted, in standard

form:

Table B

	Receptor Identifier (Codon Mutation)	Mutation Probe Sequence (5'-3') (SEQ.ID.NO.)	Selection Marker Probe Sequence (5'-3') (SEQ.ID.NO.)
20	GPR1 (F245K)	GATCTCCAGTAGGCAT <u>AAG</u> T GGACAATTCTGG (131)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAG (132)
	GPR4 (K223A)	AGAAGGCCAAGATC <u>GCG</u> CGG CTGGCCCTCA (133)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
25	GPR5 (V224K)	CGGCGCCACGCACGAAAAA GCTCATCTTC	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT

	(134)	
GPR7	GCCAAGAAGCGGGTG <u>AAG</u> TT	CTCCTTCGGTCCTCCTATCGT
(T250K)	CCTGGTGGTGGCA (135)	TGTCAGAAGT
· L	1(100)	1
GPR8	CAGGCGGAAGGTGAAAGTCC	CTCCTTCGGTCCTCCTATCGT
(T259K)	TGGTCCTCGT	TGTCAGAAGT
(1259K)	(136)	
30 GPR9	CGGCGCCTGCGGGCCAAGCG	CTCCTTCGGTCCTCCTATCGT
(M254K)	GCTGGTGGTG	TGTCAGAAGT
(IVI234K)	(137)	1
GPR9-6	CCAAGCACAAAGCCAAGAAA	CTCCTTCGGTCCTCCTATCGT
(L241K)	GTGACCATCAC	TGTCAGAAGT
(LZ41K)	(138)	
GPR10	GCGCCGGCGCACCAAATGCT	CTCCTTCGGTCCTCCTATCGT
35 (F276K)	TGCTGGTGGT	TGTCAGAAGT
33 (12761)	(139)	
GPR15	CAAAAAGCTGAAGAAATCT <u>A</u>	CTCCTTCGGTCCTCCTATCGT
(1240K)	<u>AG</u> AAGATCATCTTTATTGTCG	TGTCAGAAGT
. (2 1012)	(140)	
GPR17	CAAGACCAAGGCA <u>AAA</u> CGCA	CTCCTTCGGTCCTCCTATCGT
(V234K)	TGATCGCCAT	TGTCAGAAGT
	(141)	
40 GPR18	GTCAAGGAGAAGTCC <u>AAA</u> AG	CTCCTTCGGTCCTCCTATCGT
(I231K)	GATCATCATC	TGTCAGAAGT
(1251K)	(142)	
GPR20	CGCCGCGTGCGGGCCAAGCA	CTCCTTCGGTCCTCCTATCGT
(M240K)	GCTCCTGCTC	TGTCAGAAGT
(N1240K)	(143)	
GPR21	CCTGATAAGCGCTAT <u>AAA</u> AT	CTCCTTCGGTCCTCCTATCGT
45 (A251K)	GGTCCTGTTTCGA	TGTCAGAAGT
(A231K)	(144)	
GPR22	GAAAGACAAAAGAGAGTC <u>A</u>	CTCCTTCGGTCCTCCTATCGT
(F312K)	<u>AG</u> AGGATGTCTTTATTG	TGTCAGAAGT
(201223)	(145)	
GPR24	CGGAGAAAGAGGGTG <u>AAA</u> C	CTCCTTCGGTCCTCCTATCGT
(T304K)	GCACAGCCATCGCC	TGTCAGAAGT
	(146)	
50 GPR30	alternate approach; see below	alternate approach; see below
(L258K)		••
GPR31	AAGCTTCAGCGGGCCAAGGC	CTCCTTCGGTCCTCCTATCGT
	ACTGGTCACC	TGTCAGAAGT
(Q221K)	(147)	
GPR32	CATGCCAACCGGCCCGCGAG	ACCAGCAGCAGCCTCGCGGG
1	GCTGCTGCTGGT	CCGGTTGGCATG
5\$ (K255A)	(279)	(280)
CPP 40	CGGAAGCTGCGGGCCAAATG	CTCCTTCGGTCCTCCTATCGT
GPR40	GGTGGCCGGC	TGTCAGAAGT
(A223K)		IGICAGAAGI
CDD 41	(265) CAGAGGAGGGTG <u>AAG</u> GGGCT	CTCCTTCGGTCCTCCTATCGT
GPR41		
1	GTTGGCG	TGTCAGAAGT

(A223K)	(260)	
GPR43	(266)	
(V221K)	GGCGGCGCGAGCC <u>AAG</u> GGG CTGGCTGTGG (267)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
APJ 5 (L247K)	alternate approach; see below	alternate approach; see below
BLR1	CAGCGGCAGAAGGCAAAAA	CTCCTTCGGTCCTCCTATCGT
(V258K)	GGGTGGCCATC (148)	TGTCAGAAGT
CEPR (L258K)	CGGCAGAAGCGCAT GATCCTCGCG (149)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
10 EBI1		
(I262K)	GAGCGCAACAAGGCC <u>AAA</u> A AGGTGATCATC (150)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
EBI2 (L243K)	GGTGTAAACAAAAAGGCT <u>AA</u> <u>A</u> AACACAATTATTCTTATT (151)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
ETBR-LP2	GAGAGCCAGCTCAAGAGCAC	CTCCTTCGGTCCTCCTATCGT
15 (N358K)	CGTGGTG (152)	TGTCAGAAGT
GHSR	CCACAAGCAAACC <u>AAG</u> AAAA	CTCCTTCGGTCCTCCTATCGT
(V262K)	TGCTGGCTGT (153)	TGTCAGAAGT
GPCR-CNS	CTAGAGAGTCAGATGAAGTG	CTCCTTCGGTCCTCCTATCGT
(N491K)	TACAGTAGTGGCAC (155)	TGTCAGAAGT
20 GPR-NGA	CGGACAAAGTGAAAACT <u>A</u> A	CTCCTTCGGTCCTCCTATCGT
(I275K)	AAAGATGTTCCTCATT (156)	TGTCAGAAGT
H9a and H9b	GCTGAGGTTCGCAAT <u>AAA</u> CT	CTCCTTCGGTCCTCCTATCGT
(F236K)	AACCATGTTTGTG (157)	TGTCAGAAGT
HB954	GGGAGGCCGAGCTG <u>AAA</u> GCC	CTCCTTCGGTCCTCCTATCGT
25 (H265K)	ACCCTGCTC (158)	TGTCAGAAGT
HG38 (V765K)	GGGACTGCTCTATGAAAAA CACATTGCCCTG (268)	CATCAAGTGTATCATGTGCC AAGTACGCCC (154)
HM74	CAAGATCAAGAGAGCCAAAA	CTCCTTCCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC
(I230K)		CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
30 MIG		CTCCTTCGGTCCTCCTATCGT
(T273K)	·	TGTCAGAAGT
OGR1	GCAAGGACCAGATC <u>AAG</u> CGG	CTCCTTCGGTCCTCCTATCGT
(Q227K)	/ 	TGTCAGAAGT
Serotonin 5HT _{2A} 35 (C322K)	alternate approach; see below	alternate approach; see below
Serotonin 5HT _{2C} (S310K)	alternate approach; see below	alternate approach; see below

BRISHOUTH - MUSSISSA IN

V28	CAAGAAAGCCAAAGCC <u>AAG</u>	CTCCTTCGGTCCTCCTATCGT
(I230K)	AAACTGATCCTTCTG	TGTCAGAAGT
	(162)	

The non-endogenous human GPCRs were then sequenced and the derived and verified nucleic acid and amino acid sequences are listed in the accompanying "Sequence Listing" appendix

5 to this patent document, as summarized in Table C below:

Table C

Mutated GPCR	Nucleic Acid Sequence	Amino Acid Sequence
·	Listing	Listing
GPR1	SEQ.ID.NO.: 163	SEQ.ID.NO.: 164
(F245K)		
10 GPR4	SEQ.ID.NO.: 165	SEQ.ID.NO.: 166
(K223A)		
GPR5	SEQ.ID.NO.: 167	SEQ.ID.NO.: 168
(V224K)		
GPR7	SEQ.ID.NO.: 169	SEQ.ID.NO.: 170
15 (T250K)		
GPR8	SEQ.ID.NO.: 171	SEQ.ID.NO.: 172
(T259K)	·	
GPR9	SEQ.ID.NO.: 173	SEQ.ID.NO.: 174
(M254K)		
20 GPR9-6	SEQ.ID.NO.: 175	SEQ.ID.NO.: 176
(L241K)		
GPR10	SEQ.ID.NO.: 177	SEQ.ID.NO.: 178
(F276K)		
GPR15	SEQ.ID.NO.: 179	SEQ.ID.NO.: 180
25 (I240K)		
GPR17	SEQ.ID.NO.: 181	SEQ.ID.NO.: 182
(V234K)		
GPR18	SEQ.ID.NO.: 183	SEQ.ID.NO.: 184
(I231K)		
30 GPR20	SEQ.ID.NO.: 185	SEQ.ID.NO.: 186
(M240K)		
GPR21	SEQ.ID.NO.: 187	SEQ.ID.NO.: 188
(A251K)		
GPR22	SEQ.ID.NO.: 189	SEQ.ID.NO.: 190
35 (F312K)		
GPR24	SEQ.ID.NO.: 191	SEQ.ID.NO.: 192
(T304K))		070 77 104
GPR30	SEQ.ID.NO.: 193	SEQ.ID.NO.: 194

SEQ.ID.NO.: 195 SEQ.ID.NO.: 269 SEQ.ID.NO.: 271 SEQ.ID.NO.: 273 SEQ.ID.NO.: 275 SEQ.ID.NO.: 197 SEQ.ID.NO.: 199 SEQ.ID.NO.: 201 SEQ.ID.NO.: 203	SEQ.ID.NO.: 196 SEQ.ID.NO.: 270 SEQ.ID.NO.: 272 SEQ.ID.NO.: 274 SEQ.ID.NO.: 276 SEQ.ID.NO.: 198 SEQ.ID.NO.: 200 SEQ.ID.NO.: 202
SEQ.ID.NO.: 269 SEQ.ID.NO.: 271 SEQ.ID.NO.: 273 SEQ.ID.NO.: 275 SEQ.ID.NO.: 197 SEQ.ID.NO.: 199 SEQ.ID.NO.: 201	SEQ.ID.NO.: 270 SEQ.ID.NO.: 272 SEQ.ID.NO.: 274 SEQ.ID.NO.: 276 SEQ.ID.NO.: 198 SEQ.ID.NO.: 200
SEQ.ID.NO.: 271 SEQ.ID.NO.: 273 SEQ.ID.NO.: 275 SEQ.ID.NO.: 197 SEQ.ID.NO.: 199 SEQ.ID.NO.: 201	SEQ.ID.NO.: 272 SEQ.ID.NO.: 274 SEQ.ID.NO.: 276 SEQ.ID.NO.: 198 SEQ.ID.NO.: 200
SEQ.ID.NO.: 271 SEQ.ID.NO.: 273 SEQ.ID.NO.: 275 SEQ.ID.NO.: 197 SEQ.ID.NO.: 199 SEQ.ID.NO.: 201	SEQ.ID.NO.: 272 SEQ.ID.NO.: 274 SEQ.ID.NO.: 276 SEQ.ID.NO.: 198 SEQ.ID.NO.: 200
SEQ.ID.NO.: 273 SEQ.ID.NO.: 275 SEQ.ID.NO.: 197 SEQ.ID.NO.: 199 SEQ.ID.NO.: 201	SEQ.ID.NO.: 274 SEQ.ID.NO.: 276 SEQ.ID.NO.: 198 SEQ.ID.NO.: 200
SEQ.ID.NO.: 273 SEQ.ID.NO.: 275 SEQ.ID.NO.: 197 SEQ.ID.NO.: 199 SEQ.ID.NO.: 201	SEQ.ID.NO.: 274 SEQ.ID.NO.: 276 SEQ.ID.NO.: 198 SEQ.ID.NO.: 200
SEQ.ID.NO.: 275 SEQ.ID.NO.: 197 SEQ.ID.NO.: 199 SEQ.ID.NO.: 201	SEQ.ID.NO.: 276 SEQ.ID.NO.: 198 SEQ.ID.NO.: 200
SEQ.ID.NO.: 275 SEQ.ID.NO.: 197 SEQ.ID.NO.: 199 SEQ.ID.NO.: 201	SEQ.ID.NO.: 276 SEQ.ID.NO.: 198 SEQ.ID.NO.: 200
SEQ.ID.NO.: 197 SEQ.ID.NO.: 199 SEQ.ID.NO.: 201	SEQ.ID.NO.: 198 SEQ.ID.NO.: 200
SEQ.ID.NO.: 199 SEQ.ID.NO.: 201	SEQ.ID.NO.: 198 SEQ.ID.NO.: 200
SEQ.ID.NO.: 199 SEQ.ID.NO.: 201	SEQ.ID.NO.: 200
SEQ.ID.NO.: 201	
SEQ.ID.NO.: 201	
	SEQ.ID.NO.: 202
	SEQ.ID.NO.: 202
SEQ.ID.NO.: 203	
SEQ.ID.NO.: 203	
1	SEQ.ID.NO.: 204
SEO ID NO : 205	STO TO NO.
3EQ.ID.140 203	SEQ.ID.NO.: 206
SEO ID NO : 207	SEQ.ID.NO.: 208
52Q.12.1(0 207	32Q.1D.110 208
SEO.ID.NO.: 209	SEQ.ID.NO.: 210
SEQ.ID.NO.: 211	SEQ.ID.NO.: 212
SEQ.ID.NO.: 213	SEQ.ID.NO.: 214
SEQ.ID.NO.: 215	SEQ.ID.NO.: 216
GEO. TO 210	
SEQ.ID.NO.: 217	SEQ.ID.NO.: 218
SEO ID NO : 210	SEQ.ID.NO.: 220
3EQ.ID.140 219	SEQ.ID.NO.: 220
SEO ID NO : 277	CEO TO VIOLENCE
SEQ.ID.140.: 277	SEQ.ID.NO.: 278
SEO ID NO : 221	SEQ.ID.NO.: 222
SEQ.ID.110 221	3EQ.ID.NO.: 222
SEO.ID.NO.: 223	SEQ.ID.NO.: 224
	3-4
SEQ.ID.NO.: 225	SEQ.ID.NO.: 226
SEQ.ID.NO.: 227	SEQ.ID.NO.: 228
	1
SEQ.ID.NO.: 229	SEQ.ID.NO.: 230
(= 2.5	32Q.12.11O 230
SEO.ID.NO : 231	SEQ.ID.NO.: 232
	5LQ.ID.140 232
	SEQ.ID.NO.: 205 SEQ.ID.NO.: 207 SEQ.ID.NO.: 209 SEQ.ID.NO.: 211 SEQ.ID.NO.: 213 SEQ.ID.NO.: 215 SEQ.ID.NO.: 217 SEQ.ID.NO.: 217 SEQ.ID.NO.: 217 SEQ.ID.NO.: 221 SEQ.ID.NO.: 221 SEQ.ID.NO.: 223 SEQ.ID.NO.: 225

2. Alternate Mutation Approaches for Employment of the Proline Marker Algorithm: APJ; Serotonin 5HT_{2A}; Serotonin 5HT_{2C}; and GPR30

Although the above site-directed mutagenesis approach is particularly preferred, other approaches can be utilized to create such mutations; those skilled in the art are readily credited with selecting approaches to mutating a GPCR that fits within the particular needs of the artisan.

a. APJ

Preparation of the non-endogenous, human APJ receptor was accomplished by mutating L247K. Two oligonucleotides containing this mutation were synthesized:

- 5'- GGCTTAAGAGCATCATCGTGGTGCTGGTG-3' (SEQ.ID.NO.: 233)
- 10 5'-GTCACCACCACCACCACGATGATGCTCTTAAGCC-3' (SEQ.ID.NO.: 234)

The two oligonucleotides were annealed and used to replace the NaeI-BstEII fragment of human; endogenous APJ to generate the non-endogenous, version of human APJ.

b. Serotonin 5HT₂₄

cDNA containing the point mutation C322K was constructed by utilizing the restriction

15 enzyme site Sph I which encompasses amino acid 322. A primer containing the C322K mutation:

5'-CAAAGAAAGTACTGGGCATCGTCTTCTTCCT-3' (SEQ.ID.NO: 235)

was used along with the primer from the 3' untranslated region of the receptor:

5'-TGCTCTAGATTCCAGATAGGTGAAAA CTTG-3' (SEQ.ID.NO.: 236)

to perform PCR (under the conditions described above). The resulting PCR fragment was then

used to replace the 3' end of endogenous 5HT_{2A} cDNA through the T4 polymerase blunted Sph I site.

c. Serotonin 5HT_{2C}

20

The cDNA containing a S310K mutation was constructed by replacing the Sty I restriction fragment containing amino acid 310 with synthetic double stranded oligonucleotides that encode

the desired mutation. The sense strand sequence utilized had the following sequence:

5'-CTAGGGGCACCATGCAGGCTATCAACAATGAAAGAAAAGCTAAGAAAGTC-3'

(SEQ. ID.NO.: 237)

and the antisense strand sequence utilized had the following sequence:

5 5'-CAAGGACTTTCTTAGCTTTTCTTTCATTGTTGATAGCCTGCATGGTGCCC-3' (SEQ.

ID. NO.: 238)

d. GPR30

Prior to generating non-endogenous GPR30, several independent pCR2.1/GPR30 isolates were sequenced in their entirety in order to identify clones with no PCR-generated mutations. A clone having no mutations was digested with EcoR1 and the endogenous GPR30 cDNA fragment was transferred into the CMV-driven expression plasmid pCI-neo (Promega), by digesting pCI-Neo with EcoR1 and subcloning the EcoRI-liberated GPR30 fragment from pCR2.1/GPR30, to generate pCI/GPR30. Thereafter, the leucine at codon 258 was mutated to a lysine using a Quick-Change™ Site-Directed Mutagenesis Kit (Stratagene, #200518), according to manufacturer's instructions, and the following primers:

5'-CGGCGGCAGAAGGCGAAACGCATGATCCTCGCGGT-3' (SEQ.ID.NO.: 239) and

5'-ACCGCGAGGATCATGCGTTTCGCCTTCTGC CGCCG-3' (SEQ.ID.NO.: 240)

Example 3

Receptor (Endogenous and Mutated) Expression

20

Although a variety of cells are available to the art for the expression of proteins, it is most preferred that mammalian cells be utilized. The primary reason for this is predicated upon practicalities, *i.e.*, utilization of, *e.g.*, yeast cells for the expression of a GPCR, while possible,

WO 00/22129 PCT/US99/23938

- 57 -

introduces into the protocol a non-mammalian cell which may not (indeed, in the case of yeast, does not) include the receptor-coupling, genetic-mechanism and secretary pathways that have evolved for mammalian systems – thus, results obtained in non-mammalian cells, while of potential use, are not as preferred as that obtained from mammalian cells. Of the mammalian cells, COS-7, 293 and 293T cells are particularly preferred, although the specific mammalian cell utilized can be predicated upon the particular needs of the artisan.

Unless otherwise noted herein, the following protocol was utilized for the expression of the endogenous and non-endogenous human GPCRs. Table D lists the mammalian cell and number utilized (per 150mm plate) for GPCR expression.

10

15

Table D

Receptor Name	Mammalian Cell
(Endogenous or Non-	(Number Utilized)
Endogenous)	
GPR17	293 (2 x 10 ⁴)
GPR30	293 (4 x 10 ⁴)
APJ	COS-7 (5X10 ⁶)
ETBR-LP2	293 (1 x 10 ⁷)
	293T (1 x 10 ⁷)
GHSR	293 (1 x 10 ⁷)
	293T (1 x 10 ⁷)
MIG	293 (1 x 10 ⁷)
Serotonin 5HT _{2A}	293T (1 x 10 ⁷)
Serotonin 5HT _{2c}	293T (1 x 10 ⁷)

20

25

On day one, mammalian cells were plated out. On day two, two reaction tubes were prepared (the proportions to follow for each tube are per plate): tube A was prepared by mixing 20µg DNA (e.g., pCMV vector; pCMV vector with endogenous receptor cDNA, and pCMV vector with non-endogenous receptor cDNA.) in 1.2ml serum free DMEM (Irvine Scientific,

Irvine, CA); tube B was prepared by mixing 120µl lipofectamine (Gibco BRL) in 1.2ml serum free DMEM. Tubes A and B were then admixed by inversions (several times), followed by incubation at room temperature for 30-45min. The admixture is referred to as the "transfection mixture". Plated cells were washed with 1XPBS, followed by addition of 10ml serum free DMEM. 2.4ml of the transfection mixture was then added to the cells, followed by incubation for 4hrs at 37°C/5% CO₂. The transfection mixture was then removed by aspiration, followed by the addition of 25ml of DMEM/10% Fetal Bovine Serum. Cells were then incubated at 37°C/5% CO₂. After 72hr incubation, cells were then harvested and utilized for analysis.

1. Gi-Coupled Receptors: Co-Transfection with Gs-Coupled Receptors

In the case of GPR30, it has been determined that this receptor couples the G protein Gi. Gi is known to inhibit the enzyme adenylyl cyclase, which is necessary for catalyzing the conversion of ATP to cAMP. Thus, a non-endogenous, constitutively activated form of GPR30 would be expected to be associated with decreased levels of cAMP. Assay confirmation of a non-endogenous, constitutively activated form of GPR30 directly via measurement of decreasing levels of cAMP, while viable, can be preferably measured by cooperative use of a Gs-coupled receptor. For example, a receptor that is Gs-coupled will stimulate adenylyl cyclase, and thus will be associated with an increase in cAMP. The assignee of the present application has discovered that the orphan receptor GPR6 is an endogenous, constitutively activated GPCR. GPR6 couples to the Gs protein. Thus when co-transfected, one can readily verify that a putative GPR30-mutation leads to constitutive activation thereof: i.e., an endogenous, constitutively activated GPR6/endogenous, non-constitutively activated GPR30 cell will evidence an elevated level of cAMP when compared with an endogenous, constitutively active GPR6/non-endogenous, constitutively activated GPR30 (the latter evidencing a comparatively lower level of cAMP).

Assays that detect cAMP can be utilized to determine if a candidate compound is e.g., an inverse agonist to a Gs-associated receptor (i.e., such a compound would decrease the levels of cAMP) or a Gi-associated receptor (or a Go-associated receptor) (i.e., such a candidate compound would increase the levels of cAMP). A variety of approaches known in the art for measuring cAMP can be utilized; a preferred approach relies upon the use of anti-cAMP antibodies. Another approach, and most preferred, utilizes a whole cell second messenger reporter system assay. Promoters on genes drive the expression of the proteins that a particular gene encodes. Cyclic AMP drives gene expression by promoting the binding of a cAMP-responsive DNA binding protein or transcription factor (CREB) which then binds to the promoter at specific sites called cAMP response elements 10 and drives the expression of the gene. Reporter systems can be constructed which have a promoter containing multiple cAMP response elements before the reporter gene, e.g., \u03b3-galactosidase or luciferase. Thus, an activated receptor such as GPR6 causes the accumulation of cAMP which then activates the gene and expression of the reporter protein. Most preferably, 293 cells are cotransfected with GPR6 (or another Gs-linked receptor) and GPR30 (or another Gi-linked receptor) plasmids, preferably in a 1:1 ratio, most preferably in a 1:4 ratio. Because GPR6 is an endogenous, constitutively active receptor that stimulates the production of cAMP, GPR6 strongly activates the reporter gene and its expression. The reporter protein such as β-galactosidase or luciferase can then be detected using standard biochemical assays (Chen et al. 1995). Cotransfection of endogenous, constitutively active GPR6 with endogenous, non-constitutively active GPR30 evidences an increase in the luciferase reporter protein. Conversely, co-transfection of endogenous, constitutively active GPR6 with non-endogenous, constitutively active GPR30 evidences a drastic decrease in expression of luciferase. Several reporter plasmids are known and available in the art for measuring a second messenger assay. It is considered well within the

skilled artisan to determine an appropriate reporter plasmid for a particular gene expression based primarily upon the particular need of the artisan. Although a variety of cells are available for expression, mammalian cells are most preferred, and of these types, 293 cells are most preferred. 293 cells were transfected with the reporter plasmid pCRE-Luc/GPR6 and non-endogenous, constitutively activated GPR30 using a Mammalian Transfection™ Kit (Stratagene, #200285) CaPO₄ precipitation protocol according to the manufacturer's instructions (*see*, 28 Genomics 347 (1995) for the published endogenous GPR6 sequence). The precipitate contained 400ng reporter, 80ng CMV-expression plasmid (having a 1:4 GPR6 to endogenous GPR30 or non-endogenous GPR30 ratio) and 20ng CMV-SEAP (a transfection control plasmid encoding secreted alkaline phosphatase). 50% of the precipitate was split into 3 wells of a 96-well tissue culture dish (containing 4X10⁴ cells/well); the remaining 50% was discarded. The following morning, the media was changed. 48 hr after the start of the transfection, cells were lysed and examined for luciferase activity using a Luclite™ Kit (Packard, Cat. # 6016911) and Trilux 1450 Microbeta™ liquid scintillation and luminescence counter (Wallac) as per the vendor's instructions. The data were analyzed using GraphPad Prism 2.0a (GraphPad Software Inc.).

With respect to GPR17, which has also been determined to be Gi-linked, a modification of the foregoing approach was utilized, based upon, *inter alia*, use of another Gs-linked endogenous receptor, GPR3 (*see* 23 Genomics 609 (1994) and 24 Genomics 391 (1994)). Most preferably, 293 cells are utilized. These cells were plated-out on 96 well plates at a density of 2 x 10⁴ cells per well and were transfected using Lipofectamine Reagent (BRL) the following day according to manufacturer instructions. A DNA/lipid mixture was prepared for each 6-well transfection as follows: 260ng of plasmid DNA in 100µl of DMEM were gently mixed with 2µl of lipid in 100µl of DMEM (the 260ng of plasmid DNA consisted of 200ng of a 8xCRE-Luc

5

1747 1 18 13 25 1

10

15

reporter plasmid (see below), 50ng of pCMV comprising endogenous receptor or non-endogenous receptor or pCMV alone, and 10ng of a GPRS expression plasmid (GPRS in pcDNA3 (Invitrogen)). The 8XCRE-Luc reporter plasmid was prepared as follows: vector SRIF-β-gal was obtained by cloning the rat somatostatin promoter (-71/+51) at BgIV-HindIII site in the pßgal-Basic Vector (Clontech). Eight (8) copies of cAMP response element were obtained by PCR from an adenovirus template AdpCF126CCRE8 (see 7 Human Gene Therapy 1883 (1996)) and cloned into the SRIF- β -gal vector at the Kpn-BglV site, resulting in the 8xCRE- β -gal reporter vector. The 8xCRE-Luc reporter plasmid was generated by replacing the beta-galactosidase gene in the 8xCRE-β-gal reporter vector with the luciferase gene obtained from the pGL3-basic vector (Promega) at the HindIII-BamHI site. Following 30min. incubation at room temperature, the DNA/lipid mixture was diluted with 400 µl of DMEM and 100µl of the diluted mixture was added to each well. 100 µl of DMEM with 10% FCS were added to each well after a 4hr incubation in a cell culture incubator. The next morning the transfected cells were changed with 200 µl/well of DMEM with 10% FCS. Eight (8) hours later, the wells were changed to 100 µl /well of DMEM without phenol red, after one wash with PBS. Luciferase activity were measured the next day using the LucLite™ reporter gene assay kit (Packard) following manufacturer instructions and read on a 1450 MicroBetaTM scintillation and luminescence counter (Wallac).

Figure 4 evidences that constitutively active GPR30 inhibits GPR6-mediated activation of CRE-Luc reporter in 293 cells. Luciferase was measured at about 4.1 relative light units in the expression vector pCMV. Endogenous GPR30 expressed luciferase at about 8.5 relative light units, whereas the non-endogenous, constitutively active GPR30 (L258K), expressed luciferase at about 3.8 and 3.1 relative light units, respectively. Co-transfection of endogenous GPR6 with endogenous GPR30, at a 1:4 ratio, drastically increased luciferase

expression to about 104.1 relative light units. Co-transfection of endogenous GPR6 with non-endogenous GPR30 (L258K), at the same ratio, drastically decreased the expression, which is evident at about 18.2 and 29.5 relative light units, respectively. Similar results were observed with respect to GPR17 with respect to co-transfection with GPR3, as set forth in Figure 5.

Example 3 ASSAYS FOR DETERMINATION OF CONSTITUTIVE ACTIVITY OF NON-ENDOGENOUS GPCRS

A. Membrane Binding Assays

1. $[^{35}S]GTP\gamma S$ Assay

When a G protein-coupled receptor is in its active state, either as a result of ligand binding or constitutive activation, the receptor couples to a G protein and stimulates the release of GDP and subsequent binding of GTP to the G protein. The alpha subunit of the G protein-receptor complex acts as a GTPase and slowly hydrolyzes the GTP to GDP, at which point the receptor normally is deactivated. Constitutively activated receptors continue to exchange GDP for GTP. The non-hydrolyzable GTP analog, [35S]GTPγS, can be utilized to demonstrate enhanced binding of [35S]GTPγS to membranes expressing constitutively activated receptors. The advantage of using [35S]GTPγS binding to measure constitutive activation is that: (a) it is generically applicable to all G protein-coupled receptors; (b) it is proximal at the membrane surface making it less likely to pick-up molecules which affect the intracellular cascade.

The assay utilizes the ability of G protein coupled receptors to stimulate [35S]GTPγS binding to membranes expressing the relevant receptors. The assay can, therefore, be used in the direct identification method to screen candidate compounds to known, orphan and constitutively activated G protein-coupled receptors. The assay is generic and has application

to drug discovery at all G protein-coupled receptors.

The [35S]GTPγS assay was incubated in 20 mM HEPES and between 1 and about 20mM MgCl₂ (this amount can be adjusted for optimization of results, although 20mM is preferred) pH 7.4, binding buffer with between about 0.3 and about 1.2 nM [35S]GTPγS (this amount can be adjusted for optimization of results, although 1.2 is preferred) and 12.5 to 75 μg membrane protein (*e.g.*, COS-7 cells expressing the receptor; this amount can be adjusted for optimization, although 75μg is preferred) and 1 μM GDP (this amount can be changed for optimization) for 1 hour. Wheatgerm agglutinin beads (25 μl; Amersham) were then added and the mixture was incubated for another 30 minutes at room temperature. The tubes were then centrifuged at 1500 x g for 5 minutes at room temperature and then counted in a scintillation counter.

A less costly but equally applicable alternative has been identified which also meets the needs of large scale screening. Flash platesTM and WallacTM scintistrips may be utilized to format a high throughput [35S]GTPγS binding assay. Furthermore, using this technique, the assay can be utilized for known GPCRs to simultaneously monitor tritiated ligand binding to the receptor at the same time as monitoring the efficacy via [35S]GTPγS binding. This is possible because the Wallac beta counter can switch energy windows to look at both tritium and 35S-labeled probes. This assay may also be used to detect other types of membrane activation events resulting in receptor activation. For example, the assay may be used to monitor 32P phosphorylation of a variety of receptors (both G protein coupled and tyrosine kinase receptors). When the membranes are centrifuged to the bottom of the well, the bound [35S]GTPγS or the 32P-phosphorylated receptor will activate the scintillant which is coated of the wells. Scinti® strips (Wallac) have been used to demonstrate this principle. In addition, the assay also has utility for measuring ligand binding to receptors using radioactively labeled ligands. In a similar manner, when the radiolabeled bound

ligand is centrifuged to the bottom of the well, the scintistrip label comes into proximity with the radiolabeled ligand resulting in activation and detection.

Representative results of graph comparing Control (pCMV), Endogenous APJ and Non-Endogenous APJ, based upon the foregoing protocol, are set forth in Figure 6.

2. Adenylyl Cyclase

A Flash PlateTM Adenylyl Cyclase kit (New England Nuclear; Cat. No. SMP004A) designed for cell-based assays was modified for use with crude plasma membranes. The Flash Plate wells contain a scintillant coating which also contains a specific antibody recognizing cAMP. The cAMP generated in the wells was quantitated by a direct competition for binding of radioactive cAMP tracer to the cAMP antibody. The following serves as a brief protocol for the measurement of changes in cAMP levels in membranes that express the receptors.

Transfected cells were harvested approximately three days after transfection. Membranes were prepared by homogenization of suspended cells in buffer containing 20mM HEPES, pH 7.4 and 10mM MgCl₂. Homogenization was performed on ice using a Brinkman Polytron™ for approximately 10 seconds. The resulting homogenate was centrifuged at 49,000 X g for 15 minutes at 4°C. The resulting pellet was then resuspended in buffer containing 20mM HEPES, pH 7.4 and 0.1 mM EDTA, homogenized for 10 seconds, followed by centrifugation at 49,000 X g for 15 minutes at 4°C. The resulting pellet can be stored at -80°C until utilized. On the day of measurement, the membrane pellet was slowly thawed at room temperature, resuspended in buffer containing 20mM HEPES, pH 7.4 and 10mM MgCL₂ (these amounts can be optimized, although the values listed herein are prefereed), to yield a final protein concentration of 0.60mg/ml (the resuspended membranes were placed on ice until use).

cAMP standards and Detection Buffer (comprising 2 μ Ci of tracer [125I cAMP (100 μ l] to

BNSDOCID - WO DOSSISSA IA-

5

15

The rest of

20

11 ml Detection Buffer) were prepared and maintained in accordance with the manufacturer's instructions. Assay Buffer was prepared fresh for screening and contained 20mM HEPES, pH 7.4, 10mM MgCl₂, 20mM (Sigma), 0.1 units/ml creatine phosphokinase (Sigma), 50 µM GTP (Sigma), and 0.2 mM ATP (Sigma); Assay Buffer can be stored on ice until utilized. The assay was initiated by addition of 50ul of assay buffer followed by addition of 50ul of membrane suspension to the NEN Flash Plate. The resultant assay mixture is incubated for 60 minutes at room temperature followed by addition of 100ul of detection buffer. Plates are then incubated an additional 2-4 hours followed by counting in a Wallac MicroBeta scintillation counter. Values of cAMP/well are extrapolated from a standard cAMP curve which is contained within each assay plate. The foregoing assay was utilized with respect to analysis of MIG.

B. Reporter-Based Assays

1. CREB Reporter Assay (Gs-associated receptors)

A method to detect Gs stimulation depends on the known property of the transcription factor CREB, which is activated in a cAMP-dependent manner. A PathDetect CREB trans-Reporting System (Stratagene, Catalogue # 219010) was utilized to assay for Gs coupled activity in 293 or 293T cells. Cells were transfected with the plasmids components of this above system and the indicated expression plasmid encoding endogenous or mutant receptor using a Mammalian Transfection Kit (Stratagene, Catalogue #200285) according to the manufacurer's instructions. Briefly, 400 ng pFR-Luc (luciferase reporter plasmid containing Gal4 recognition sequences), 40 ng pFA2-CREB (Gal4-CREB fusion protein containing the Gal4 DNA-binding domain), 80 ng CMV-receptor expression plasmid (comprising the receptor) and 20 ng CMV-SEAP (secreted alkaline phosphatase expression plasmid; alkaline phosphatase activity is measured in the media of transfected cells to control for variations in

transfection efficiency between samples) were combined in a calcium phosphate precipitate as per the Kit's instructions. Half of the precipitate was equally distributed over 3 wells in a 96-well plate, kept on the cells overnight, and replaced with fresh medium the following morning. Forty-eight (48) hr after the start of the transfection, cells were treated and assayed for luciferase activity as set forth with resepct to the GPR30 system, above. This assay was used with respect to GHSR.

2. AP1 reporter assay (Gq-associated receptors)

Ae method to detect Gq stimulation depends on the known property of Gq-dependent phospholipase C to cause the activation of genes containing AP1 elements in their promoter. A Pathdetect AP-1 cis-Reporting System (Stratagene, Catalogue # 219073) was utilized following the protocl set forth above with respect to the CREB reporter assay, except that the components of the calcium phosphate precipitate were 410 ng pAP1-Luc, 80 ng receptor expression plasmid, and 20 ng CMV-SEAP. This assay was used with respect to ETBR-LP2

C. Intracellular IP3 Accumulation Assay

On day 1, cells comprising the serotonin receptors (endogenous and mutated) were plated onto 24 well plates, usually 1x10⁵ cells/well. On day 2 cells were transfected by firstly mixing 0.25 ug DNA in 50 ul serumfree DMEM/well and 2 ul lipofectamine in 50 µl serumfree DMEM/well. The solutions were gently mixed and incubated for 15-30 min at room temperature. Cells were washed with 0.5 ml PBS and 400 µl of serum free media was mixed with the transfection media and added to the cells. The cells were then incubated for 3-4 hrs at 37°C/5%CO₂ and then the transfection media was removed and replaced with 1ml/well of regular growth media. On day 3 the cells were labeled with ³H-myo-inositol. Briefly, the media was removed the cells were washed with 0.5 ml PBS. Then 0.5 ml inositol-free/serumfree media (GIBCO BRL) was added/well with 0.25 µCi of ³H-myo-inositol/well

10

15

20

and the cells were incubated for 16-18 hrs o/n at 37°C/5%CO₂. On Day 4 the cells were washed with 0.5 ml PBS and 0.45 ml of assay medium was added containing inositolfree/serum free media 10 µM pargyline 10 mM lithium chloride or 0.4 ml of assay medium and 50 ul of 10x ketanserin (ket) to final concentration of 10µM. The cells were then incubated for 30 min at 37°C. The cells were then washed with 0.5 ml PBSand 200 ul of fresh/icecold stop solution (1M KOH; 18 mM Na-borate; 3.8 mM EDTA) was added/well. The solution was kept on ice for 5-10 min or until cells were lysed and then neutralized by 200 µl of fresh/ice cold neutralization sol. (7.5 % HCL). The lysate was then transferred into 1.5 ml eppendorf tubes and 1 ml of chloroform/methanol (1:2) was added/tube. The solution was vortexed for 15 sec and the upper phase was applied to a Biorad AG1-X8 anion exchange resin (100-200 mesh). Firstly, the resin was washed with water at 1:1.25 W/V and 0.9 ml of upper phase was loaded onto the column. The column was washed with 10 mls of 5 mM myo-inositol and 10 ml of 5 mM Na-borate/60mM Na-formate. The inositol tris phosphates were eluted into scintillation vials containing 10 ml of scintillation cocktail with 2 ml of 0.1 M formic acid/ 1 M ammonium formate. The columns were regenerated by washing with 10 ml of 0.1 M formic acid/3M ammonium formate and rinsed twice with dd H₂O and stored at 4°C in water.

Figure 7 provides an illustration of IP3 production from the human 5-HT_{2A} receptor that incorporates the C322K mutation. While these results evidence that the Proline Mutation Algorithm approach constitutively activates this receptor, for purposes of using such a receptor for screening for identification of potential therapeutics, a more robust difference would be preferred. However, because the activated receptor can be utilized for understanding and elucidating the role of constitutive activation and for the identification of compounds that

can be further examined, we believe that this difference is itself useful in differentiating between the endogenous and non-endogenous versions of the human $5HT_{2A}$ receptor.

D. Result Summary

The results for the GPCRs tested are set forth in Table E where the Per-Cent Increase indicates the percentage difference in results observed for the non-endogenous GPCR as compared to the endogenous GPCR; these values are followed by parenthetical indications as to the type of assay utilized. Additionally, the assay sytem utilized is parenthetically listed (and, in cases where different Host Cells were used, both are listed). As these results indicate, a variety of assays can be utilized to determine constitutive activity of the non-endogenous versions of the human GPCRs.

Those skilled in the art, based upon the foregoing and with reference to information available to the art, are creditied with theability to selelect and/ot maximize a particular assay approach that suites the particular needs of theinvestigator.

Table E

20	

Receptor Identifier	Per-Cent Difference
(Codon Mutation)	
GPR17	74.5
(V234K)	(CRE-Luc)
GPR30	71.6
(L258K)	(CREB)
APJ	49.0
(L247K)	(GTP _γ S)
ETBR-LP2	48.4(AP1-Luc - 293)
(N358K)	61.1(AP1-Luc - 293T)

GHSR	58.9(CREB - 293)
(V262K)	35.6(CREB - 293T)

MIG	39 (cAMP)
(I230K)	A Company of the Comp
Serotonin 5HT _{2A}	33.2 (IP ₃)
(C322K)	
Serotonin 5HT _{2C}	39.1(IP ₃)
(S310K)	

Example 6

Tissue Distribution of Endogenous Orphan GPCRs

Using a commercially available human-tissue dot-blot format, endogenous orphan GPCRs were probed for a determination of the areas where such receptors are localized. Except as indicate below, the entire receptor cDNA (radiolabelled) was used as the probe: radiolabeled probe was generated using the complete receptor cDNA (excised from the vector) using a Prime-It II™ Random Primer Labeling Kit (Stratagene, #300385), according to manufacturer's instructions. A human RNA Master Blot™ (Clontech, #7770-1) was hybridized with the GPCR radiolabeled probe and washed under stringent conditions according manufacturer's instructions. The blot was exposed to Kodak BioMax Autoradiography film overnight at -80°C.

Representative dot-blot format results are presented in Figure 8 for GPR1 (8A), GPR30 (8B), and APJ (8C), with results being summarized for all receptors in Table F

Table F

GPCR	Tissue Distribution
	(highest levels, relative to other tissues in
	the dot-blot)
GPR1	Placenta, Ovary, Adrenal

5

20

	GPR4	Decade thickers in Tleast Town Advanced
	GPR4	Broad; highest in Heart, Lung, Adrenal,
		Thyroid, Spinal Cord
	GPR5	Placenta, Thymus, Fetal Thymus
		Lesser levels in spleen, fetal spleen
	GPR7	Liver, Spleen, Spinal Cord, Placenta
	GPR8	No expression detected
	GPR9-6	Thymus, Fetal Thymus
		Lesser levels in Small Intestine
	GPR18	Spleen, Lymph Node, Fetal Spleen, Testis
	GPR20	Broad
	GPR21	Broad; very low abundance
	GPR22	Heart, Fetal Heart
	• • •	Lesser levels in Brain
	GPR30	Stomach
	GPR31	Broad
	BLR1	Spleen
	CEPR	Stomach, Liver, Thyroid, Putamen
	EBI1	Pancreas
		Lesser levels in Lymphoid Tissues
	EBI2	Lymphoid Tissues, Aorta, Lung, Spinal Cord
	ETBR-LP2	Broad; Brain Tissue
	GPCR-CNS	Brain
	•	Lesser levels in Testis, Placenta
	GPR-NGA	Pituitary
		Lesser levels in Brain
	H9	Pituitary
	HB954	Aorta, Cerebellum
n.		Lesser levels in most other tissues
	HM74	Spleen, Leukocytes, Bone marrow, Mammary
		Glands, Lung, Trachea
	MIG	Low levels in Kidney, Liver, Pancreas, Lung,
,		Spleen
	ORG1	Pituitary, Stomach, Placenta
	V28	Brain, Spleen, Peripheral Leukocytes

Based upon the foregoing information, it is noted that human GPCRs can also be assessed for distribution in diseased tissue; comparative assessments between "normal" and diseased tissue can then be utilized to determine the potential for over-expression or under-expression of a particular receptor in a diseased state. In those circumstances where it is desirable to utilize the non-endogenous versions of the human GPCRs for the purpose of screening to directly identify

candidate compounds of potential therapeutic relevance, it is noted that inverse agonists are useful in the treatment of diseases and disorders where a particular human GPCR is over-expressed, whereas agonists or partial agonists are useful in the treatment of diseases and disorders where a particular human GPCR is under-expressed.

As desired, more detailed, cellular localization of the recepotrs, using techniques well-known to those in the art (e.g., in-situ hybridization) can be utilized to identify particular cells within these tissues where the receptor of interest is expressed.

5

10

It is intended that each of the patents, applications, and printed publications mentioned in this patent document be hereby incorporated by reference in their entirety.

As those skilled in the art will appreciate, numerous changes and modifications may be made to the preferred embodiments of the invention without departing from the spirit of the invention. It is intended that all such variations fall within the scope of the invention.

Although a variety of expression vectors are available to those in the art, for purposes of utilization for both the endogenous and non-endogenous human GPCRs, it is most preferred that the vector utilized be pCMV. This vector has been deposited with the American Type Culture Collection (ATCC) on October 13, 1998 (10801 University Blvd., Manassas, VA 20110-2209 USA) under the provisions of the Budapest Treaty for the International Recognition of the Deposit of Microorganisms for the Purpose of patent Procedure. The vector was tested by the ATCC on _______, 1998 and determined to be viable on _______, 1998. The ATCC has assigned the following deposit number to pCMV: ______.

CLAIMS

What is claimed is:

1. A constitutively active, non-endogenous version of an endogenous human orphan G protein-coupled receptor (GPCR) comprising the following amino acid residues (carboxy-terminus to amino-terminus orientation) transversing the transmembrane-6 (TM6) and intracellular loop-3 (IC3) regions of the non-endogenous GPCR:

PI AA₁₅ X

wherein:

- (1) P¹ is an amino acid residue located within the TM6 region of the nonendogenous GPCR, where P¹ is selected from the group consisting of (i) the endogenous orphan GPCR proline residue, and (ii) a nonendogenous amino acid residue other than proline;
- (2) AA₁₅ are 15 amino acid residues selected from the group consisting of (a) the 15 endogenous amino acid residues of the endogenous orphan GPCR, (b) 15 non-endogenous amino acid residues, and (c) a combination of 15 amino acid residues, the combination comprising at least one endogenous amino acid residue of the endogenous orphan GPCR and at least one non-endogenous amino acid residue, excepting that none of the 15 endogenous amino acid residues that are positioned within the TM6 region of the GPCR is proline; and
- (2) X is a non-endogenous amino acid residue located within the IC3 region of said non-endogenous GPCR.
- 2. The non-endogenous human GPCR of claim 1 wherein P¹ is the endogenous proline

10

15

5

15

residue.

- 3. The non-endogenous human GPCR of claim1 wherein P¹ is a non-endogenous amino acid residue other than a proline residue.
- 4. The non-endogenous human GPCR of claim 1 wherein AA₁₅ are the 15 endogenous amino acid residues of the endogenous GPCR.
- 5. The non-endogenous human GPCR of claim 1 wherein X is selected from the group consisting of lysine, hisitidine, arganine and alanine residues, excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X is selected from the group consisting of histidine, arginine and alanine.
- The non-endogenous human GPCR of claim 1 wherein X is a lysine residue, excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X is an amino acid other than lysine.
 - 7. The non-endogenous human GPCR of claim 4 wherein X is a lysine residue, excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X is an amino acid other than lysine.
 - 8. The non-endogenous, human GPCR of claim 1 wherein P¹ is a proline residue and X is a lysine residue, excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X is an amino acid other than lysine.
 - 9. A host cell comprising the non-endogenous human GPCR of claim 1.
- 20 10. The material of claim 9 wherein said host cell is of mammalian origin.
 - 11. The non-endogenous human GPCR of claim 1 in a purified and isolated form.
 - 12. A nucleic acid sequence encoding a constitutively active, non-endogenous version of an endogenous human orphan G protein-coupled receptor (GPCR) comprising the following

nucleic acid sequence region transversing the transmembrane-6 (TM6) and intracellular loop-3 (IC3) regions of the orphan GPCR:

3'-P^{codon} (AA-codon)₁₅ X_{codon}-5'

wherein:

5

10

- (1) P^{codon} is a nucleic acid encoding region within the TM6 region of the non-endogenous GPCR, where P^{codon} encodes an amino acid selected from the group consisting of (i) the endogenous GPCR proline residue, and (ii) a non-endogenous amino acid residue other than proline;
- (2) (AA-codon)₁₅ are 15 codons encoding 15 amino acid residues selected from the group consisting of (a) the 15 endogenous amino acid residues of the endogenous orphan GPCR, (b) 15 non-endogenous amino acid residues, and (c) a combination of 15 amino acid residues, the combination comprising at least one endogenous amino acid residue of the endogenous orphan GPCR and at least one non-endogenous amino acid residue, excepting that none of the 15

15

(3) X_{codon} is a nucleic acid encoding region residue located within the IC3 region of said non-endogenous human GPCR, where X_{codon} encodes a non-endogenous amino acid.

endogenous amino acid residues that are positioned within the TM6

20

13. The nucleic acid sequence of claim 12 wherein P^{codon} encodes an endogenous proline residue.

region of the orphan GPCR is proline; and

14. The nucleic acid sequence of claim 12 wherein P^{codon} encodes a non-endogenous

5

10

amino acid residue other than a proline residue.

- 15. The nucleic acid sequence of claim 12 wherein X_{codon} encodes a non-endogenous amino acid selected from the group consisting of lysine, histidine, arginine and alanine, excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X_{codon} encodes an amino acid selected from the group consisiting of histidine, arginine and alanine.
- 16. The nucleic acid sequence of claim 13 wherein X_{codon} encodes a non-endogenous lysine amino acid excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X_{codon} encodes an amino acid selected from the group consisiting of histidine, arginine and alanine.
- 17. The nucleic acid sequence of claim 12 wherein X_{codon} is selected from the group consisting of AAA, AAG, GCA, GCG, GCC and GCU.
- 18. The nucleic acid sequence of claim 12 wherein X_{codon} is selected from the group consisting of AAA and AAG.
- 15 19. The nucleic acid sequence of claim 12 wherein P^{codon} is selected from the group consisting of CCA, CCC, CCG and CCU, and X_{codon} is selected from the group consisting of AAA and AAG.
 - 20. A vector comprising the nucleic acid sequence of claim 12.
 - 21. A plasmid comprising the nucleic acid sequence of claim 12.
- 20 22. A host cell comprising the nucleic acid sequence of claim 21.
 - 23. The nucleic acid sequence of claim 12 in a purified and isolated form.
 - 24. A method for selecting for alteration an endogenous amino acid residue within the third intracellular loop of a human G protein-coupled receptor ("GPCR"), said receptor

comprising a transmembrane 6 region and an intracellular loop 3 region, which endogenous amino acid, when altered to a non-endogenous amino acid, constitutively activates said human GPCR, comprising the following steps:

- (a) identifying an endogenous proline residue within the transmembrane 6 region of a human GPCR;
 - (b) identifying, by moving in a direction of the carboxy-terminus region of said GPCR towards the amino-terminus region of said GPCR, the endogenous, 16th amino acid residue from said proline residue;
- (c) altering the endogenous residue of step (b) to a non-endogenous amino acid residue to create a non-endogenous version of an endogenous human GPCR; and
 - (d) determining whether the non-endogenous human GPCR of step (c) is constitutively active.
- The method of claim 24 wherein the amino acid residue that is two residues from said
 proline residue in the transmembrane 6 region, in a carboxy-terminus to aminoterminus direction, is tryptophan.
 - A constitutively active, non-endogenous human GPCR produced by the process of claim 24.
- 27. A constitutively active, non-endogenous human GPCR produced by the process of claim 25.
 - 28. An algorithmic approach for creating a non-endogenous, constitutively active version of an endogenous human G protein coupled receptor (GPCR), said endogenous GPCR comprising a transmembrane 6 region and an intracellular loop 3 region, the

5

algorithmic approach comprising the steps of:

- (a) selecting an endogenous human GPCR comprising a proline residue in the transmembrane-6 region;
- (b) identifying, by counting 16 amino acid residues from the proline residue of step (a), in a carboxy-terminus to amino-terminus direction, an endogenous amino acid residue;
- altering the identified amino acid residue of step (b) to a non-endogenous amino acid residue to create a non-endogenous version of the endogenous human GPCR; and
- 10 (d) determining if the non-endogenous version of the endogenous human GPCR of step (c) is constitutively active.
 - 29. The algorithmic approach of claim 28 wherein the amino acid residue that is two residues from said proline residue in the transmembrane 6 region, in a carboxy-terminus to amino-terminus direction, is tryptophan.
- 15 30. A constitutively active, non-endogenous human GPCR produced by the algorithmic approach of claim 28.
 - 31. A constitutively active, non-endogenous human GPCR produced by the algorithmic approach of claim 29.
- 32. A method for directly identifying a compound selected from the group consisting of inverse agonists, agonists and partial agonists to a non-endogenous, constitutively activated human G protein coupled receptor, said receptor comprising a transmembrane-6 region and an intracellular loop-3 region, comprising the steps of:
 - (a) selecting an endogenous human GPCR;

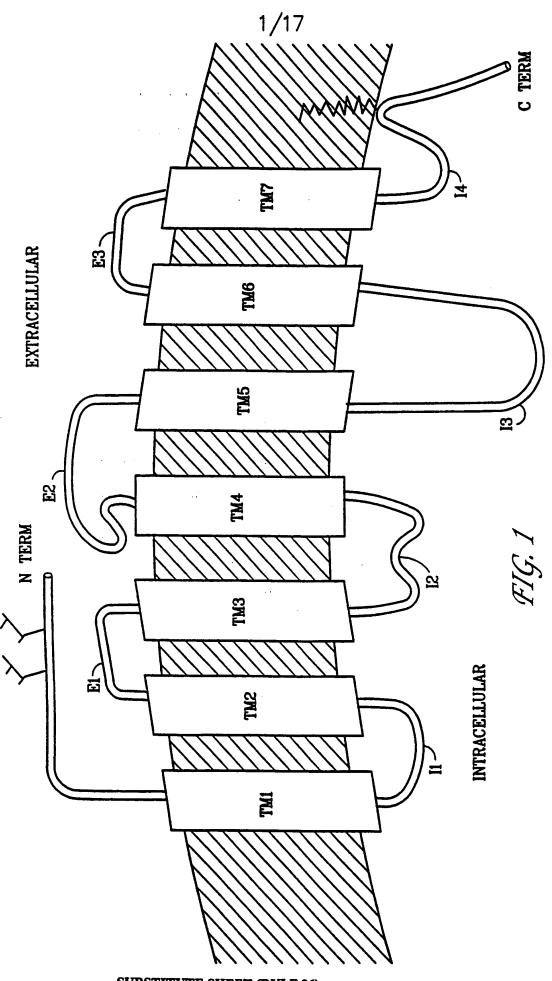
- (b) identifying a proline residue within the transmembrane-6 region of the GPCR of step (a);
- (c) identifying, in a carboxy-terminus to amino-terminus direction, the endogenous, 16th amino acid residue from the proline residue of step (b);
- 5 (d) altering the endogenous amino acid of step (c) to a non-endogenous amino acid;
 - (e) confirming that the non-endogenous GPCR of step (d) is constitutively active;
 - (f) contacting a candidate compound with the non-endogenous, constitutivelyactivated GPCR of step (e); and
- 10 (g) determining, by measurement of the compound efficacy at said contacted receptor, whether said compound is an inverse agonist, agonist or partial agonist of said receptor.
 - 33. The method of claim 32 wherein the non-endogenous amino acid of step (d) is lysine.
 - 34. A compound directly identified by the method of claim 32.
- 15 35. The method of claim 32 wherein the directly identified compound is an inverse agonist.
 - 36. The method of claim 32 wherein the directly identified compound is an agonist.--
 - 37. The method of claim 32 wherein the directly identified compound is a partial agonist.
 - 38. A composition comprising the inverse agonist of claim 35.
- 20 39. A composition comprising the agonist of claim 36.
 - 40. A composition comprising the partial agonist of claim 37.
 - 41. A method for directly identifying an inverse agonist to a non-endogenous,

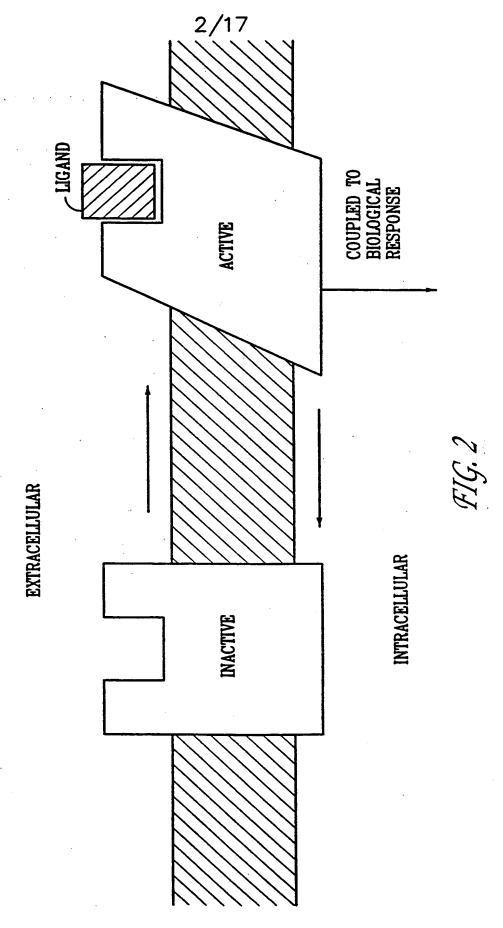
WO 00/22129 PCT/US99/23938

constitutively activated human G protein coupled receptor ("GPCR"), said GPCR comprising a transmembrane-6 region and an intracellular loop-3 region, comprising the steps of:

(a) selecting an endogenous human GPCR;

- (b) identifying a proline residue within the transmembrane-6 region of the GPCR of step (a);
 - (c) identifying, in a carboxy-terminus to amino-terminus direction, the endogenous, 16th amino acid residue from the proline residue of step (b);
 - (d) altering the endogenous amino acid of step (c) to a non-endogenous lysine residue;
 - (e) confirming that the non-endogenous GPCR of step (d) is constitutively active;
- (f) contacting a candidate compound with the non-endogenous, constitutivelyactivated GPCR of step (e); and
 - (g) determining, by measurement of the compound efficacy at said contacted receptor, whether said compound is an inverse agonist of said receptor.
 - 42. An inverse agonist directly identified by the method of claim 37.
- 15 43. A composition comprising an inverse agonist of claim 38.





SUBSTITUTE SHEET (RULE 26)

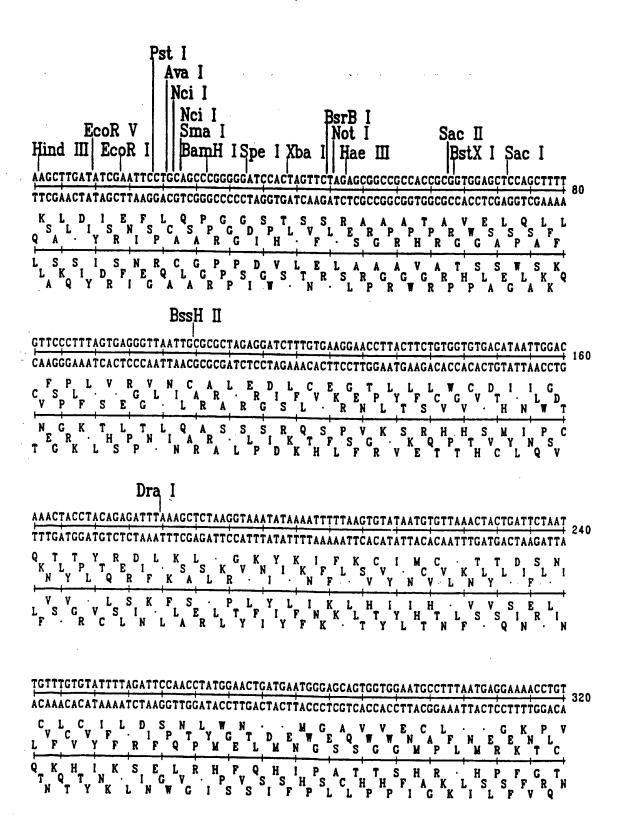
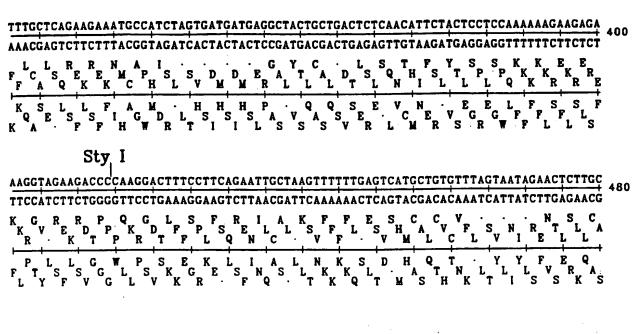
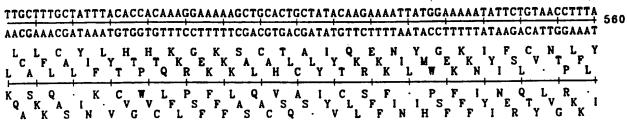


FIG. 3A





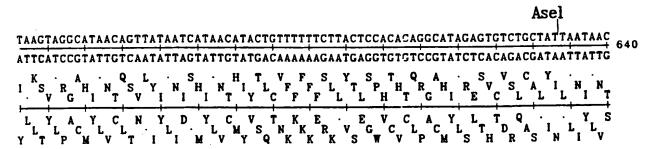




FIG. 3B

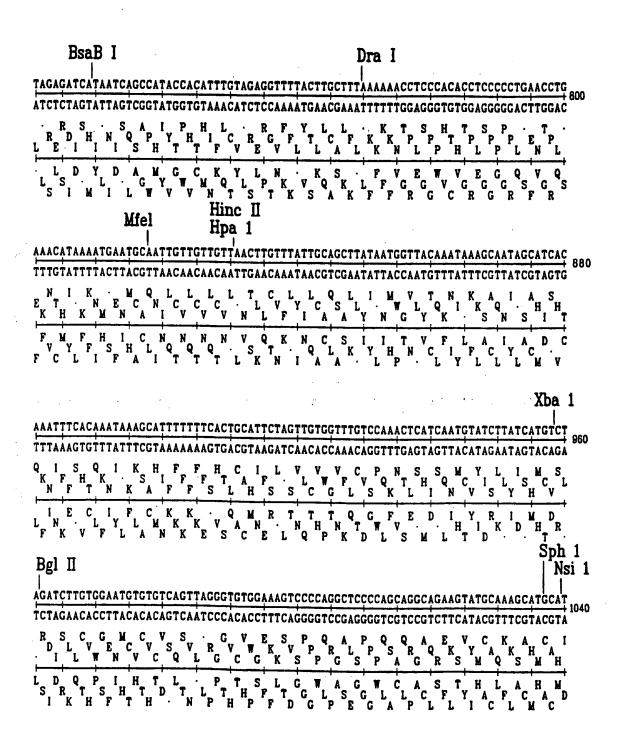


FIG. 3C

6/17

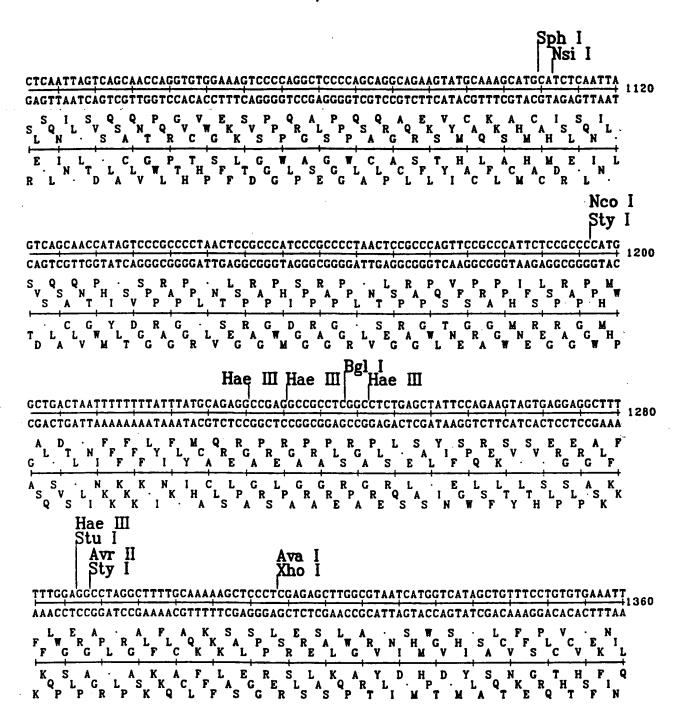


FIG. 3D

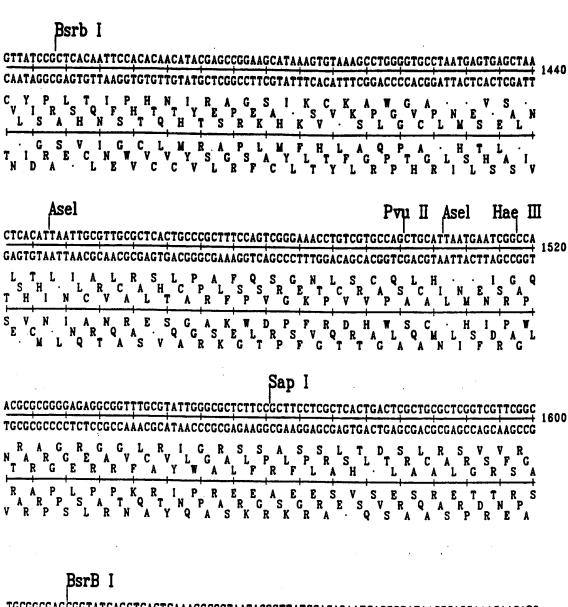


FIG. 3E

8/17

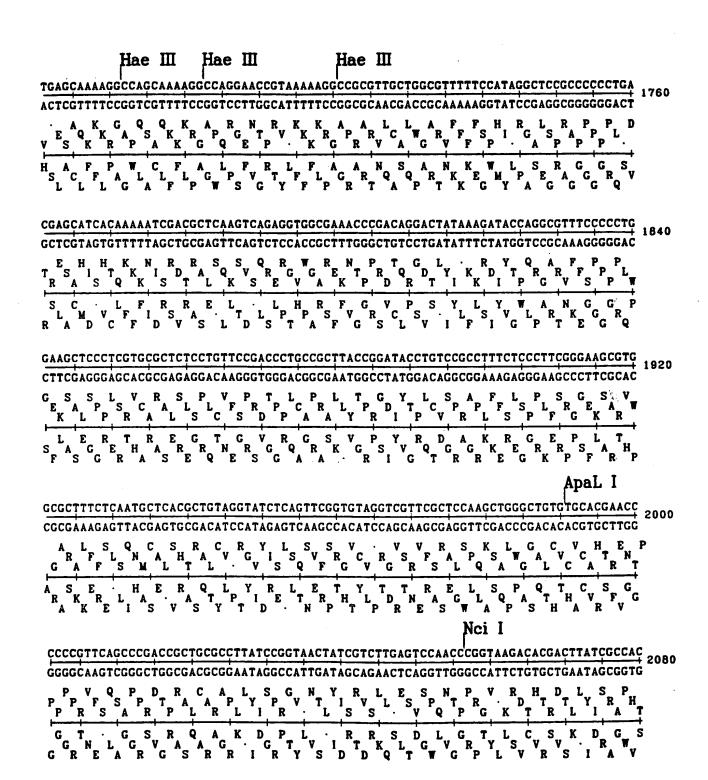


FIG. 3F

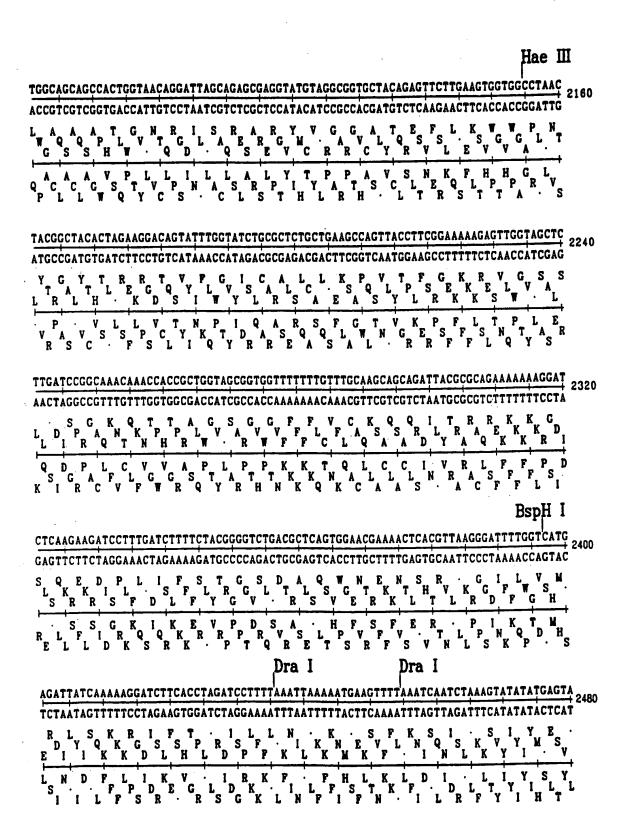


FIG. 3G

10/17

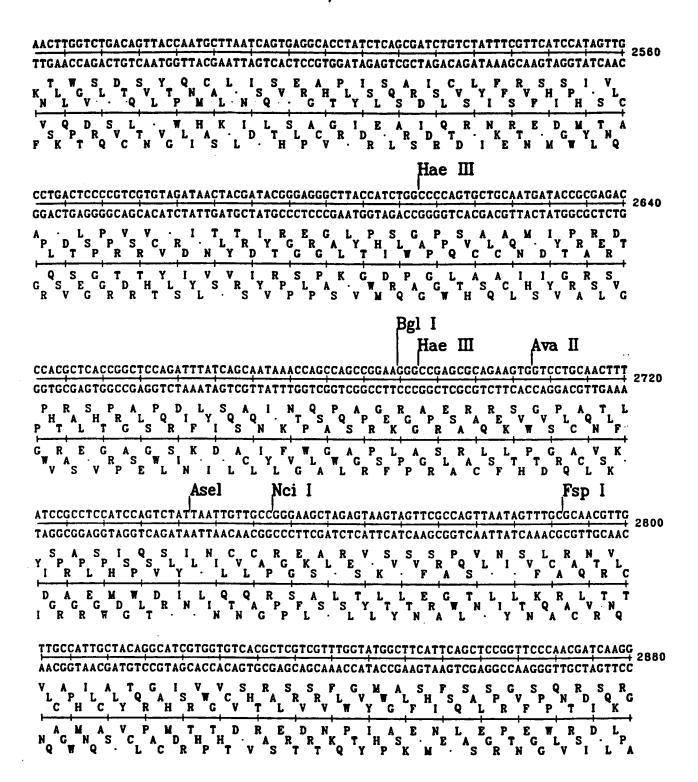


FIG. 3H

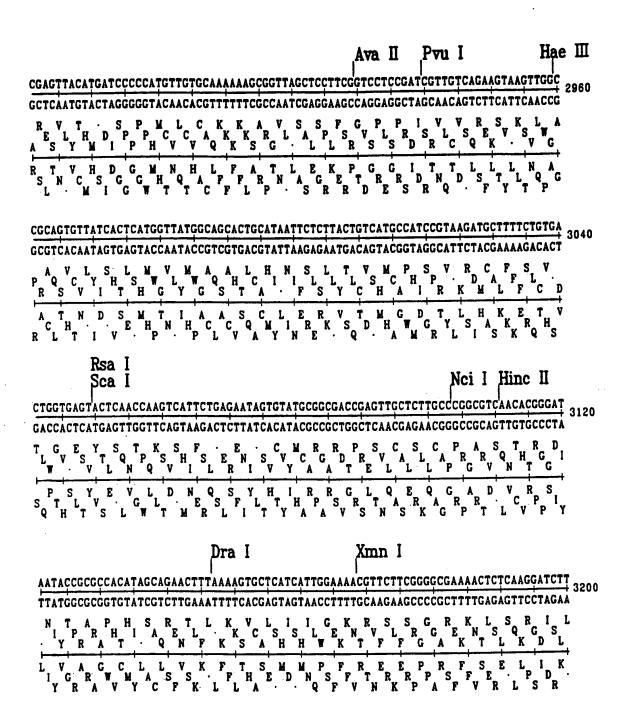


FIG. 31

12/17

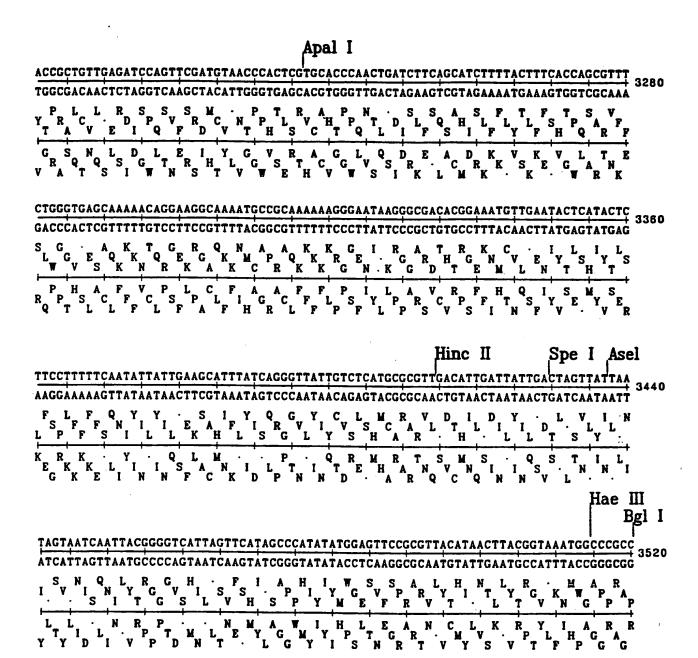


FIG. 3J

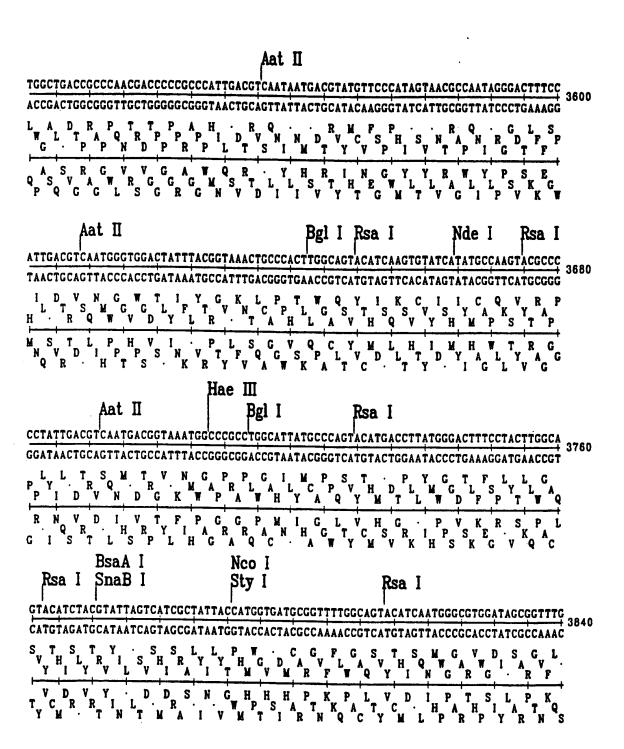
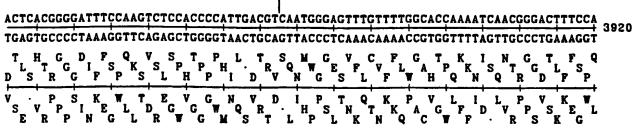


FIG. 3K

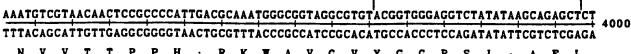
Sac I

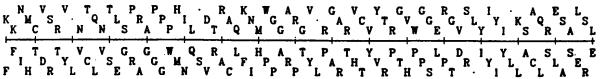
14/17





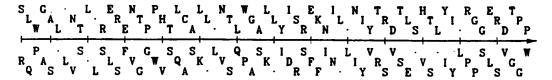
Rsa I

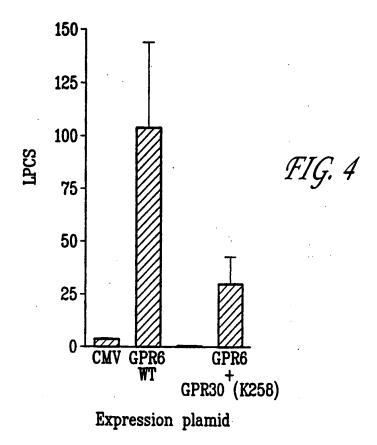


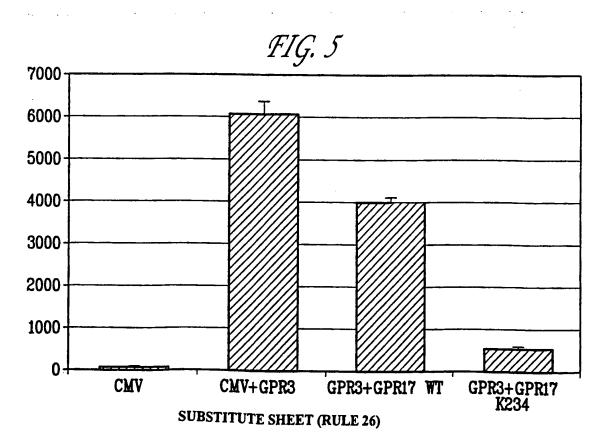


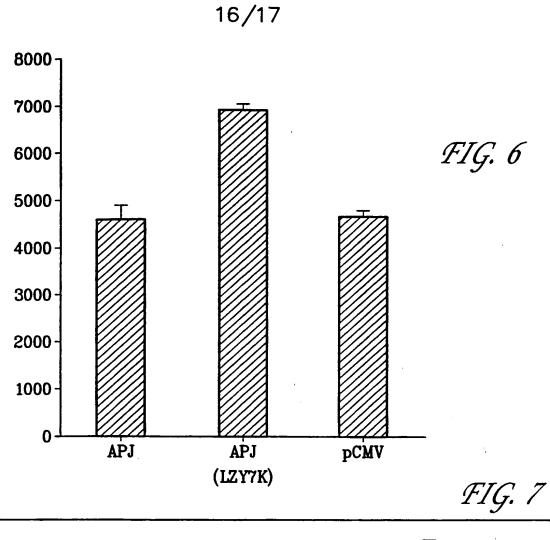
Rsa I

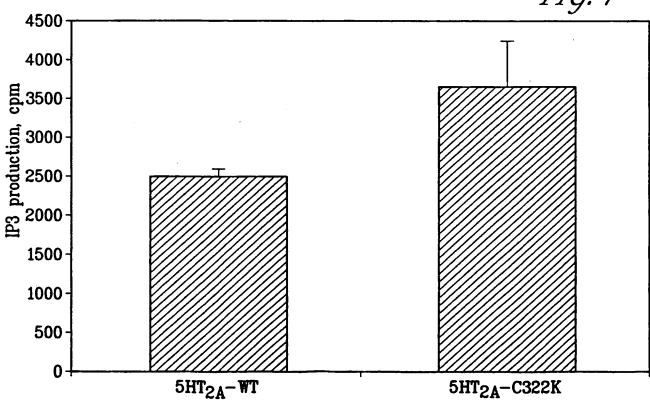
CTGGCTAACTAGAGAACCCACTGCTTAACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCC











SUBSTITUTE SHEET (RULE 26)

17/17

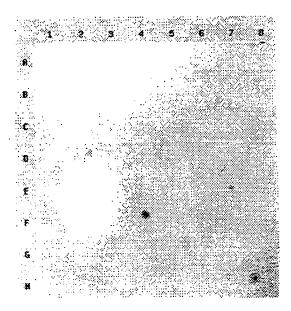
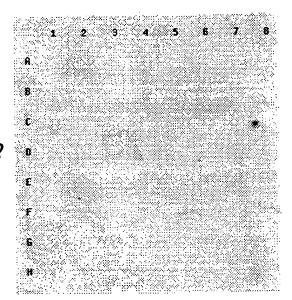


FIG. 8A

FIG. 8B



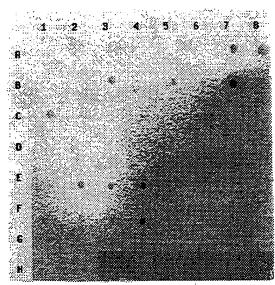


FIG. 8C

SUBSTITUTE SHEET (RULE 26)

WO 00/22129 PCT/US99/23938

1

SEQUENCE LISTING

(1) GENERAL INFORMATION: (i) APPLICANT: Behan, Dominic P. Chalmers, Derek T. 5 Liaw, Chen W. (ii) TITLE OF INVENTION: Non-Endogenous, Constitutively Activated Human G Protein-Coupled Orphan Receptors (iii) NUMBER OF SEQUENCES: 280 10 (iv) CORRESPONDENCE ADDRESS: (A) ADDRESSEE: Arena Pharmaceuticals, Inc. 6166 Nancy Ridge Drive (B) STREET: (C) CITY: San Diego 15 (D) STATE: CA (E) COUNTRY: USA (F) ZIP: (v) COMPUTER READABLE FORM: (A) MEDIUM TYPE: Floppy disk 20 (B) COMPUTER: IBM PC compatible (C) OPERATING SYSTEM: PC-DOS/MS-DOS (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (vi) CURRENT APPLICATION DATA: (A) APPLICATION NUMBER: US 25 (B) FILING DATE: (C) CLASSIFICATION: (viii) ATTORNEY/AGENT INFORMATION: (A) NAME: Burgoon, Richard P. (B) REGISTRATION NUMBER: 34,787 30 (ix) TELECOMMUNICATION INFORMATION: (A) TELEPHONE: (619)453-7200 (B) TELEFAX: (619)453-7210 (2) INFORMATION FOR SEQ ID NO:1: (i) SEQUENCE CHARACTERISTICS: 35 (A) LENGTH: 1068 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (genomic) 40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

	TATTACTCTC	TGGAGTCTGA	TTTGGAGGAG	AAAGTCCAGC	TGGGAGTTGT	TCACTGGGTC	120
	TCCCTGGTGT	TATATTGTTT	GGCTTTTGTT	CTGGGAATTC	CAGGAAATGC	CATCGTCATT	180
	TGGTTCACGG	GGCTCAAGTG	GAAGAAGACA	GTCACCACTC	TGTGGTTCCT	CAATCTACCC	240
	ATTGCGGATT	TCATTTTTCT	TCTCTTTCTG	CCCCTGTACA	TCTCCTATGT	GGCCATGAAT	300
5	TTCCACTGGC	CCTTTGGCAT	CTGGCTGTGC	AAAGCCAATT	CCTTCACTGC	CCAGTTGAAC	360
	ATGTTTGCCA	GTGTTTTTT	CCTGACAGTG	ATCAGCCTGG	ACCACTATAT	CCACTTGATC	420
	CATCCTGTCT	TATCTCATCG	GCATCGAACC	CTCAAGAACT	CTCTGATTGT	CATTATATTC	480
	ATCTGGCTTT	TGGCTTCTCT	AATTGGCGGT	CCTGCCCTGT	ACTTCCGGGA	CACTGTGGAG	540
	TTCAATAATC	ATACTCTTTG	CTATAACAAT	TTTCAGAAGC	ATGATCCTGA	CCTCACTTTG	600
10	ATCAGGCACC	ATGTTCTGAC	TTGGGTGAAA	TTTATCATTG	GCTATCTCTT	CCCTTTGCTA	660
	ACAATGAGTA	TTTGCTACTT	GTGTCTCATC	TTCAAGGTGA	AGAAGCGAAC	AGTCCTGATC	720
	TCCAGTAGGC	ATTTCTGGAC	AATTCTGGTT	GTGGTTGTGG	CCTTTGTGGT	TTGCTGGACT	780
	CCTTATCACC	TGTTTAGCAT	TTGGGAGCTC	ACCATTCACC	ACAATAGCTA	TTCCCACCAT	840
	GTGATGCAGG	CTGGAATCCC	CCTCTCCACT	GGTTTGGCAT	TCCTCAATAG	TTGCTTGAAC	900
15	CCCATCCTTT	ATGTCCTAAT	TAGTAAGAAG	TTCCAAGCTC	GCTTCCGGTC	CTCAGTTGCT	960
	GAGATACTCA	AGTACACACT	GTGGGAAGTC	AGCTGTTCTG	GCACAGTGAG	TGAACAGCTC	1020
	AGGAACTCAG	AAACCAAGAA	TCTGTGTCTC	CTGGAAACAG	CTCAATAA		1068

(3) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 355 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein
- 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Glu Asp Leu Glu Glu Thr Leu Phe Glu Glu Phe Glu Asn Tyr Ser 1 5 10 15

Tyr Asp Leu Asp Tyr Tyr Ser Leu Glu Ser Asp Leu Glu Glu Lys Val 20 25 30

30 Gln Leu Gly Val Val His Trp Val Ser Leu Val Leu Tyr Cys Leu Ala 35 40 45

	Phe	Val 50	Leu	Gly	Ile	Pro	Gly 55	Asn	Ala	Ile	Val	Ile 60	Trp	Phe	Thr	Gly
	Leu 65	Lys	Trp	Lys	Lys	Thr 70	Val	Thr	Thr	Leu	Trp 75	Phe	Leu	Asn	Leu	Ala 80
5	Ile	Ala	Asp	Phe	Ile 85	Phe	Leu	Leu	Phe	Leu 90	Pro	Leu	Tyr	Ile	Ser 95	Tyr
	Val	Ala	Met	Asn 100	Phe	His	Trp	Pro	Phe 105	Gly	Ile	Trp	Leu	Cys 110	Lys	Ala
10	Asn	Ser	Phe 115	Thr	Ala	Gln	Leu	Asn 120	Met	Phe	Ala	Ser	Val 125	Phe	Phe	Leu
	Thr	Val 130	Ile	Ser	Leu	Asp	His 135	Tyr	Ile	His	Leu	Ile 140	His	Pro	Val	Leu
	Ser 145	His	Arg	His	Arg	Thr 150	Leu	Lys	Asn	Ser	Leu 155	Ile	Val	Ile	Ile	Phe 160
15	Ile	Trp	Leu	Leu	Ala 165	Ser	Leu	Ile	Gly	Gly 170	Pro	Ala	Leu	Tyr	Phe 175	Arg
	Asp	Thr	Val	Glu 180	Phe	Asn	Asn	His	Thr 185	Leu	Cys	Tyr	Asn	Asn 190	Phe	Gln
20	Lys	His	Asp 195	Pro	Asp	Leu	Thr	Leu 200	Ile	Arg	His	His	Val 205	Leu	Thr	Trp
•	Val	Lys 210	Phe	Ile	Ile	Gly	Tyr 215	Leu	Phe	Pro	Leu	Leu 220	Thr	Met	Ser	Ile
	Cys 225	Tyr	Leu	Cys	Leu	Ile 230	Phe	Lys	Val	Lys	Lys 235	Arg	Thr	Val	Leu	Ile 240
25	Ser	Ser	Arg	His	Phe 245	Trp	Thr	Ile	Leu	Val 250	Val	Val	Val	Ala	Phe 255	Val
	Val	Cys	Trp	Thr 260	Pro	Tyr	His	Leu	Phe 265	Ser	Ile	Trp	Glu	Leu 270	Thr	Ile
30	His	His	Asn 275	Ser	Tyr	Ser	His	His 280	Val	Met	Gln	Ala	Gly 285	Ile	Pro	Leu
	Ser	Thr 290	Gly	Leu	Ala	Phe	Leu 295	Asn	Ser	Cys	Leu	Asn 300	Pro	Ile	Leu	Tyr
	Val 305	Leu	Ile	Ser	Lys	Lys 310	Phe	Gln	Ala	Arg	Phe 315	Arg	Ser	Ser	Val	Ala 320
35	Glu	Ile	Leu	Lys	Tyr 325	Thr	Leu	Trp	Glu	Val 330	Ser	Cys	Ser	Gly	Thr 335	Val
	Ser	Glu	Gln	Leu	Arq	Asn	Ser	Glu	Thr	Lys	Asn	Leu	Cys	Leu	Leu	Glu

..... 340 345 350

Thr Ala Gln 355

(4) INFORMATION FOR SEQ ID NO:3:

5 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1089 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATGGGCAACC ACACGTGGGA GGGCTGCCAC GTGGACTCGC GCGTGGACCA CCTCTTTCCG 60 CCATCCCTCT ACATCTTTGT CATCGGCGTG GGGCTGCCCA CCAACTGCCT GGCTCTGTGG 120 GCGGCCTACC GCCAGGTGCA ACAGCGCAAC GAGCTGGGCG TCTACCTGAT GAACCTCAGC 180 15 ATCGCCGACC TGCTGTACAT CTGCACGCTG CCGCTGTGGG TGGACTACTT CCTGCACCAC 240 GACAACTGGA TCCACGGCCC CGGGTCCTGC AAGCTCTTTG GGTTCATCTT CTACACCAAT 300 ATCTACATCA GCATCGCCTT CCTGTGCTGC ATCTCGGTGG ACCGCTACCT GGCTGTGGCC 360 CACCCACTCC GCTTCGCCCG CCTGCGCCGC GTCAAGACCG CCGTGGCCGT GAGCTCCGTG 420 GTCTGGGCCA CGGAGCTGGG CGCCAACTCG GCGCCCCTGT TCCATGACGA GCTCTTCCGA 480 20 GACCGCTACA ACCACACCTT CTGCTTTGAG AAGTTCCCCA TGGAAGGCTG GGTGGCCTGG 540 ATGAACCTCT ATCGGGTGTT CGTGGGCTTC CTCTTCCCGT GGGCGCTCAT GCTGCTGTCG 600 TACCGGGGCA TCCTGCGGGC CGTGCGGGGC AGCGTGTCCA CCGAGCGCCA GGAGAAGGCC 660 AAGATCAAGC GGCTGGCCCT CAGCCTCATC GCCATCGTGC TGGTCTGCTT TGCGCCCTAT CACGTGCTCT TGCTGTCCCG CAGCGCCATC TACCTGGGCC GCCCCTGGGA CTGCGGCTTC 25 GAGGAGCGCG TCTTTTCTGC ATACCACAGC TCACTGGCTT TCACCAGCCT CAACTGTGTG GCGGACCCCA TCCTCTACTG CCTGGTCAAC GAGGGCGCCC GCAGCGATGT GGCCAAGGCC 900 CTGCACAACC TGCTCCGCTT TCTGGCCAGC GACAAGCCCC AGGAGATGGC CAATGCCTCG 960 CTCACCCTGG AGACCCCACT CACCTCCAAG AGGAACAGCA CAGCCAAAGC CATGACTGGC 1020 AGCTGGGCGG CCACTCCGCC TTCCCAGGGG GACCAGGTGC AGCTGAAGAT GCTGCCGCCA 1080 30 GCACAATGA 1089

WO 00/22129 PCT/US99/23938

5

(5) INFORMATION FOR SEQ ID NO:4:

5

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 362 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

	(1) 31	SQUEI	NCE 1	JESCI	XIPI.	LOIN:	SEQ	י עד	NO:4	•					
10	Met 1	Gly	Asn	His	Thr 5	Trp	Glu	Gly	Cys	His 10	Val	Asp	Ser	Arg	Val 15	Asp
	His	Leu	Phe	Pro 20	Pro	Ser	Leu	Tyr	Ile 25	Phe	Val	Ile	Gly	Val 30	Gly	Leu
	Pro	Thr	Asn 35	Cys	Leu	Ala	Leu	Trp 40	Ala	Ala	Tyr	Arg	Gln 45	Val	Gln	Gln
15	Arg	Asn 50	Glu	Leu	Gly	Val	Tyr 55	Leu	Met	Asn	Leu	Ser 60	Ile	Ala	Asp	Leu 🖁
	Leu 65	Tyr	Ile	Cys	Thr	Leu 70	Pro	Leu	Trp		Asp 75	Tyr	Phe	Leu	His	His 80
20	Asp	Asn	Trp		His 85	Gly	Pro	Gly	Ser	Cys 90	Lys	Leu	Phe	Gly	Phe 95	Ile
	Phe	Tyr	Thr	Asn 100		Tyr	Ile	Ser	Ile 105	Ala	Phe	Leu	Cys	Cys 110	Ile	Ser
	Val	Asp	Arg 115	Tyr	Leu	Ala	Val	Ala 120	His	Pro	Leu	Arg	Phe 125	Ala	Arg	Leu
25	Arg	Arg 130	Val	Lys	Thr	Ala	Val 135	Ala	Val	Ser	Ser	Val 140	Val	Trp	Ala	Thr
	Glu 145	Leu	Gly	Ala	Asn	Ser 150	Ala	Pro	Leu	Phe	His 155	Asp	Glu	Leu	Phe	Arg 160
30	Asp	Arg	туr	Asn	His 165	Thr	Phe	Cys	Phe	Glu 170	Lys	Phe	Pro	Met	Glu 175	Gly
	Trp	Val	Ala	Trp 180	Met	Asn	Leu	Tyr	Arg 185	Val	Phe	Val	Gly	Phe 190	Leu	Phe
	Pro	Trp	Ala 195	Leu	Met	Leu	Leu	Ser 200	Tyr	Arg	Gly	Ile	Leu 205	Arg	Ala	Val
35	Arg	Gly 210	Ser	Val	Ser	Thr	Glu 215	Arg	Gln	Glu	Lys	Ala 220	Lys	Ile	Lys	Arg
	_	_	_		_					_			_	_	•	

Leu Ala Leu Ser Leu Ile Ala Ile Val Leu Val Cys Phe Ala Pro Tyr

	005
	225 230 235 240
	His Val Leu Leu Ser Arg Ser Ala Ile Tyr Leu Gly Arg Pro Trp 245 250 255
5	Asp Cys Gly Phe Glu Glu Arg Val Phe Ser Ala Tyr His Ser Ser Leu 260 265 270
	Ala Phe Thr Ser Leu Asn Cys Val Ala Asp Pro Ile Leu Tyr Cys Leu 275 280 285
	Val Asn Glu Gly Ala Arg Ser Asp Val Ala Lys Ala Leu His Asn Leu 290 295 300
10	Leu Arg Phe Leu Ala Ser Asp Lys Pro Gln Glu Met Ala Asn Ala Ser 305 310 315 320
	Leu Thr Leu Glu Thr Pro Leu Thr Ser Lys Arg Asn Ser Thr Ala Lys 325 330 335
15	Ala Met Thr Gly Ser Trp Ala Ala Thr Pro Pro Ser Gln Gly Asp Gln 340 345 350
	Val Gln Leu Lys Met Leu Pro Pro Ala Gln 355 360
	(6) INFORMATION FOR SEQ ID NO:5:
20	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (genomic)
25	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:
	TATGAATTCA GATGCTCTAA ACGTCCCTGC 3
	(7) INFORMATION FOR SEQ ID NO:6:
30	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (genomic)
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:
35	TCCGGATCCA CCTGCACCTG CGCCTGCACC 3
	(0) INFORMATION FOR CEO ID NO. 7.

5

4	(;)	CROTTENCE	CHARACTERISTICS:

- (A) LENGTH: 1002 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

	ATGGAGTCCT	CAGGCAACCC	AGAGAGCACC	ACCTTTTTTT	ACTATGACCT	TCAGAGCCAG	60
	CCGTGTGAGA	ACCAGGCCTG	GGTCTTTGCT	ACCCTCGCCA	CCACTGTCCT	GTACTGCCTG	120
10	GTGTTTCTCC	TCAGCCTAGT	GGGCAACAGC	CTGGTCCTGT	GGGTCCTGGT	GAAGTATGAG	180
	AGCCTGGAGT	CCCTCACCAA	CATCTTCATC	CTCAACCTGT	GCCTCTCAGA	CCTGGTGTTC	240
	GCCTGCTTGT	TGCCTGTGTG	GATCTCCCCA	TACCACTGGG	GCTGGGTGCT	GGGAGACTTC	300
	CTCTGCAAAC	TCCTCAATAT	GATCTTCTCC	ATCAGCCTCT	ACAGCAGCAT	CTTCTTCCTG	360
	ACCATCATGA	CCATCCACCG	CTACCTGTCG	GTAGTGAGCC	CCCTCTCCAC	CCTGCGCGTC	420
15	CCCACCCTCC	GCTGCCGGGT	GCTGGTGACC	ATGGCTGTGT	GGGTAGCCAG	CATCCTGTCC	480
	TCCATCCTCG	ACACCATCTT	CCACAAGGTG	CTTTCTTCGG	GCTGTGATTA	TTCCGAACTC	540
	ACGTGGTACC	TCACCTCCGT	CTACCAGCAC	AACCTCTTCT	TCCTGCTGTC	CCTGGGGATT	. 600
	ATCCTGTTCT	GCTACGTGGA	GATCCTCAGG	ACCCTGTTCC	GCTCACGCTC	CAAGCGGCGC	660
	CACCGCACGG	TCAAGCTCAT	CTTCGCCATC	GTGGTGGCCT	ACTTCCTCAG	CTGGGGTCCC	720
20	TACAACTTCA	CCCTGTTTCT	GCAGACGCTG	TTTCGGACCC	AGATCATCCG	GAGCTGCGAG	780
	GCCAAACAGC	AGCTAGAATA	CGCCCTGCTC	ATCTGCCGCA	ACCTCGCCTT	CTCCCACTGC	840
	TGCTTTAACC	CGGTGCTCTA	TGTCTTCGTG	GGGGTCAAGT	TCCGCACACA	CCTGAAACAT	900
	GTTCTCCGGC	AGTTCTGGTT	CTGCCGGCTG	CAGGCACCCA	GCCCAGCCTC	GATCCCCCAC	960
	ጥ ርርርርጥርርጥር	CCTTCGCCTA	TGAGGGCGCC	TCCTTCTACT	GA		1002

25 (9) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 333 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
- 30 (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

	Met 1	Glu	Ser	Ser	Gly 5	Asn	Pro	Glu	Ser	Thr 10	Thr	Phe	Phe	Tyr	Tyr 15	Asp
	Leu	Gln	Ser	Gln 20	Pro	Cys	Glu	Asn	Gln 25	Ala	Trp	Val	Phe	Ala 30	Thr	Leu
5	Ala	Thr	Thr 35	Val	Leu	Tyr	Cys	Leu 40	Val	Phe	Leu	Leu	Ser 45	Leu	Val	Gly
	Asn	Ser 50	Leu	Val	Leu	Trp	Val 55	Leu	Val	Lys	Tyr	Glu 60	Ser	Leu	Glu	Ser
10	Leu 65	Thr	Asn	Ile	Phe	Ile 70	Leu	Asn	Leu	Cys	Leu 75	Ser	Asp	Leu	Val	Phe 80
	Ala	Cys	Leu	Leu	Pro 85	Val	Trp	Ile	Ser	Pro 90	Tyr	His	Trp	Gly	Trp 95	Val
	Leu	Gly	Asp	Phe 100	Leu	Cys	Lys	Leu	Leu 105	Asn	Met	Ile	Phe	Ser 110	lle	Ser
15	Leu	Tyr	Ser 115	Ser	Ile	Phe	Phe	Leu 120	Thr	Ile	Met	Thr	Ile 125	His	Arg	Tyr
	Leu	Ser 130	Val	Val	Ser	Pro	Leu 135	Ser	Thr	Leu	Arg	Val 140	Pro	Thr	Leu	Arg
20	Cys 145	Arg	Val	Leu	Val	Thr 150	Met	Ala	Val	Trp	Val 155	Ala	Ser	Ile	Leu	Ser 160
	Ser	Ile	Leu	Asp	Thr 165	Ile	Phe	His	Lys	Val 170	Leu	Ser	Ser	Gly	Cys 175	Asp
	Tyr	Ser	Glu	Leu 180	Thr	Trp	Tyr	Leu	Thr 185	Ser	Val	Tyr	Gln	His 190	Asn	Leu
25	Phe	Phe	Leu 195	Leu	Ser	Leu	Gly	Ile 200	Ile	Leu	Phe	Cys	Tyr 205	Val	Glu	Ile
	Leu	Arg 210	Thr	Leu	Phe	Arg	Ser 215	Arg	Ser	Lys	Arg	Arg 220	His	Arg	Thr	Val
30	Lys 225	Leu	Ile	Phe	Ala	Ile 230	Val	Val	Ala	Tyr	Phe 235	Leu	Ser	Trp	Gly	Pro 240
	Tyr	Asn	Phe	Thr	Leu 245	Phe	Leu	Gln	Thr	Leu 250	Phe	Arg	Thr	Gln	Ile 255	Ile
	Arg	Ser	Cys	Glu 260	Ala	Lys	Gln	Gln	Leu 265	Glu	Tyr	Alạ	Leu	Leu 270	Ile	Cys
35	Arg	Asn	Leu 275	Ala	Phe	Ser	His	Cys 280	Cys	Phe	Asn	Pro	Val 285	Leu	Tyr	Val

PCT/US99/23938

	Phe Val Gly Val Lys Phe Arg Thr His Leu Lys His Val Leu Arg Glr 290 295 300	
٠	Phe Trp Phe Cys Arg Leu Gln Ala Pro Ser Pro Ala Ser Ile Pro His 305 310 315 320	
5	Ser Pro Gly Ala Phe Ala Tyr Glu Gly Ala Ser Phe Tyr 325 330	
	(10) INFORMATION FOR SEQ ID NO:9:	
10	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:	
15	GCAAGCTTGG GGGACGCCAG GTCGCCGGCT	30
	(11) INFORMATION FOR SEQ ID NO:10:	
20	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	·
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:	
	GCGGATCCGG ACGCTGGGGG AGTCAGGCTG C	31
25	(12) INFORMATION FOR SEQ ID NO:11:	
30	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 987 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:	
	ATGGACAACG CCTCGTTCTC GGAGCCCTGG CCCGCCAACG CATCGGGCCC GGACCCGGCG	60
	CTGAGCTGCT CCAACGCGTC GACTCTGGCG CCGCTGCCGG CGCCGCTGGC GGTGGCTGTA	120
35	CCAGTTGTCT ACGCGGTGAT CTGCGCCGTG GGTCTGGCGG GCAACTCCGC CGTGCTGTAC	180

	GTGTTGCTGC	GGGCGCCCG	CATGAAGACC	GTCACCAACC	TGTTCATCCT	CAACCTGGCC	240
	ATCGCCGACG	AGCTCTTCAC	GCTGGTGCTG	CCCATCAACA	TCGCCGACTT	CCTGCTGCGG	300
	CAGTGGCCCT	TCGGGGAGCT	CATGTGCAAG	CTCATCGTGG	CTATCGACCA	GTACAACACC	360
	TTCTCCAGCC	TCTACTTCCT	CACCGTCATG	AGCGCCGACC	GCTACCTGGT	GGTGTTGGCC	420
5	ACTGCGGAGT	CGCGCCGGGT	GGCCGGCCGC	ACCTACAGCG	CCGCGCGCGC	GGTGAGCCTG	480
	GCCGTGTGGG	GGATCGTCAC	ACTCGTCGTG	CTGCCCTTCG	CAGTCTTCGC	CCGGCTAGAC	540
	GACGAGCAGG	GCCGGCGCCA	GTGCGTGCTA	GTCTTTCCGC	AGCCCGAGGC	CTTCTGGTGG	600
	CGCGCGAGCC	GCCTCTACAC	GCTCGTGCTG	GGCTTCGCCA	TCCCCGTGTC	CACCATCTGT	660
	GTCCTCTATA	CCACCCTGCT	GTGCCGGCTG	CATGCCATGC	GGCTGGACAG	CCACGCCAAG	720
10	GCCCTGGAGC	GCGCCAAGAA	GCGGGTGACC	TTCCTGGTGG	TGGCAATCCT	GGCGGTGTGC	780
	CTCCTCTGCT	GGACGCCCTA	CCACCTGAGC	ACCGTGGTGG	CGCTCACCAC	CGACCTCCCG	840
	CAGACGCCGC	TGGTCATCGC	TATCTCCTAC	TTCATCACCA	GCCTGACGTA	CGCCAACAGC	900
	TGCCTCAACC	CCTTCCTCTA	CGCCTTCCTG	GACGCCAGCT	TCCGCAGGAA	CCTCCGCCAG	960
	CTGATAACTT	GCCGCGCGGC	AGCCTGA				987

15 (13) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 328 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met Asp Asn Ala Ser Phe Ser Glu Pro Trp Pro Ala Asn Ala Ser Gly
1 5 10 15

25 Pro Asp Pro Ala Leu Ser Cys Ser Asn Ala Ser Thr Leu Ala Pro Leu 20 25 30

Pro Ala Pro Leu Ala Val Ala Val Pro Val Val Tyr Ala Val Ile Cys 35 40 45

Ala Val Gly Leu Ala Gly Asn Ser Ala Val Leu Tyr Val Leu Leu Arg 50 55 60

Ala Pro Arg Met Lys Thr Val Thr Asn Leu Phe Ile Leu Asn Leu Ala 65 70 75 80

		Ile	Ala	Asp	Glu	Leu 85	Phe	Thr	Leu	Val	Leu 90	Pro	Ile	Asn	Ile	Ala 95	Ası
		Phe	Leu	Leu	Arg 100	Gln	Trp	Pro	Phe	Gly 105		Leu	Met	Cys	Lys 110	Leu	Ile
5		Val	Ala	Ile 115	Asp	Gln	Tyr	Asn	Thr 120	Phe	Ser	Ser	Leu	Tyr 125	Phe	Leu	Thr
		Val	Met 130	Ser	Ala	Asp	Arg	Tyr 135	Leu	Val	Val	Leu	Ala 140	Thr	Ala	Glu	Ser
10		Arg 145	Arg	Val	Ala	Gly	Arg 150	Thr	Tyr	Ser	Ala	Ala 155	Arg	Ala	Val	Ser	Leu 160
		Ala	Val	Trp	Gly	Ile 165	Val	Thr	Leu	Val	Val 170	Leu	Pro	Phe	Ala	Val 175	Phe
		Ala	Arg	Leu	Asp 180	Asp	Glu	Gln	Gly	Arg 185	Arg	Gln	Cys	Val	Leu 190	Val	Phe
15		Pro	Gln	Pro 195	Glu	Ala	Phe	Trp	Trp 200	Arg	Ala	Ser	Arg	Leu 205	Tyr	Thr	Leu
		Val	Leu 210	Gly	Phe	Ala	Ile	Pro 215	Val	Ser	Thr	Ile	Cys 220	Val	Leu	Tyr	Thr
20		Thr 225	Leu	Leu	Cys	Arg	Leu 230	His	Ala	Met	Arg	Leu 235	Asp	Ser	His	Ala	Lys 240
		Ala	Leu	Glu	Arg	Ala 245	Lys	Lys	Arg	Val	Thr 250	Phe	Leu	Val	Val	Ala 255	Ile
		Leu	Ala	Val	Cys 260	Leu	Leu	Cys	Trp	Thr 265	Pro	Tyr	His	Leu	Ser 270	Thr	Val
25		Val		Leu 275	Thr	Thr	Asp	Leu	Pro 280	Gln	Thr	Pro	Leu	Val 285	Ile	Ala	Ile
			290					295		Tyr			300				
30		Phe 305	Leu	Tyr	Ala	Phe	Leu 310	Asp	Ala	Ser	Phe	Arg 315	Arg	Asn	Leu		Gln 320
		Leu				325											
	(14)	INFO	RMAT.	ION	FOR	SEQ	ID N	0:13	:								

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:	
	CGGAATTCGT CAACGGTCCC AGCTACAATG	30
	(15) INFORMATION FOR SEQ ID NO:14:	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
10	(ii) MOLECULE TYPE: DNA (genomic) (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:	
	ATGGATCCCA GGCCCTTCAG CACCGCAATA T	31
	(16) INFORMATION FOR SEQ ID NO:15:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1002 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:	
	ATGCAGGCCG CTGGGCACCC AGAGCCCCTT GACAGCAGGG GCTCCTTCTC CCTCCCCACG	60
	ATGGGTGCCA ACGTCTCTCA GGACAATGGC ACTGGCCACA ATGCCACCTT CTCCGAGCCA	120
	CTGCCGTTCC TCTATGTGCT CCTGCCCGCC GTGTACTCCG GGATCTGTGC TGTGGGGCTG	180
	ACTGGCAACA CGGCCGTCAT CCTTGTAATC CTAAGGGCGC CCAAGATGAA GACGGTGACC	240
25	AACGTGTTCA TCCTGAACCT GGCCGTCGCC GACGGGCTCT TCACGCTGGT ACTGCCCGTC	300
	AACATCGCGG AGCACCTGCT GCAGTACTGG CCCTTCGGGG AGCTGCTCTG CAAGCTGGTG	360
	CTGGCCGTCG ACCACTACAA CATCTTCTCC AGCATCTACT TCCTAGCCGT GATGAGCGTG	420
	GACCGATACC TGGTGGTGCT GGCCACCGTG AGGTCCCGCC ACATGCCCTG GCGCACCTAC	480
	CGGGGGGCGA AGGTCGCCAG CCTGTGTGTC TGGCTGGGCG TCACGGTCCT GGTTCTGCCC	540
30	TTCTTCTCTT TCGCTGGCGT CTACAGCAAC GAGCTGCAGG TCCCAAGCTG TGGGCTGAGC	600
	TTCCCGTGGC CCGAGCGGGT CTGGTTCAAG GCCAGCCGTG TCTACACTTT GGTCCTGGGC	660
	TTCGTGCTGC CCGTGTGCAC CATCTGTGTG CTCTACACAG ACCTCCTGCG CAGGCTGCGG	720

DESCRIPTION OF THE PROPERTY OF

GCCGTGCGGC	TCCGC	ICTGG A	CCAAG	GCT CTA	GGCAA	.GG C	CAGO	GGG2	AA G	GTGA	CCGT	780		
CTGGTCCTCG	TCGTG	CTGGC C	STGTGC	CTC CTC	TGCTG	GA C	GCCC	CTTC	CA C	CTGG	CCTC	r 840		
GTCGTGGCCC	TGACC	aceca co	CTGCCC	CAG ACC	CCACT	GG T	CATO	CAGT	AT G	TCCT	ACGT	900		
ATCACCAGCC	TCACG	TACGC C	ACTCG	IGC CTG	AACCC	CT T	CCTC	TAC	GC C	TTTC	raga:	r 960		
GACAACTTCC	GGAAG	AACTT CO	GCAGC2	ATA TTG	CGGTG	CT G	A					1002		
(17) INFOR	MATION	FOR SE	ID NO	0:16:										
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 333 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: (D) TOPOLOGY: not relevant (ii) MOLECULE TYPE: protein (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16: 														
	_			_										
Met G 1	ln Ala	Ala Gly	His F	Pro Glu		Leu 10	Asp	Ser	Arg	Gly	Ser 15	Phe		
Ser L	eu Pro	Thr Met	Gly A	Ala Asn	Val : 25	Ser	Gln	Asp	Asn	Gly 30	Thr	Gly		
His A	sn Ala 35	Thr Phe	e Ser (Glu Pro 40	Leu :	Pro	Phe	Leu	Tyr 45	Val	Leu	Leu		
	la Val	Tyr Sei	_	Ile Cys 55	Ala	Val	Gly	Leu 60	Thr	Gly	Asn	Thr		
Ala V 65	al Ile	Leu Val	. Ile I 70	Leu Arg	Ala		Lys 75	Met	Lys	Thr	Val	Thr 80		
Asn V	al Phe	Ile Let 85	a Asn I	Leu Ala	•	Ala 90	Asp	Gly	Leu	Phe	Thr 95	Leu		
Val L	eu Pro	Val Asr 100	lle A	Ala Glu	His :	Leu	Leu	Gln	Tyr	Trp 110	Pro	Phe		
Gly G	lu Leu 115	Leu Cys	Lys I	Leu Val 120	Leu :	Ala	Val	Asp	His 125	Tyr	Asn	Ile		
	er Ser 30	Ile Tyr		Leu Ala 135	Val i	Met	Ser	Val 140	Asp	Arg	Tyr	Leu		
Val V 145	al Leu	Ala Thi	Val <i>I</i> 150	Arg Ser	Arg 1		Met 155	Pro	Trp	Arg	Thr	Tyr 160		
Arg G	ly Ala	Lys Val		Ser Leu		Val 170	Trp	Leu	Gly	Val	Thr 175	Val		

	•	
	Leu Val Leu Pro Phe Phe Ser Phe Ala Gly Val Tyr Ser Asn Glu Leu 180 185 190	
	Gln Val Pro Ser Cys Gly Leu Ser Phe Pro Trp Pro Glu Arg Val Trp 195 200 205	
5	Phe Lys Ala Ser Arg Val Tyr Thr Leu Val Leu Gly Phe Val Leu Pro 210 215 220	
	Val Cys Thr Ile Cys Val Leu Tyr Thr Asp Leu Leu Arg Arg Leu Arg 225 230 235 240	
10	Ala Val Arg Leu Arg Ser Gly Ala Lys Ala Leu Gly Lys Ala Arg Arg 255	
	Lys Val Thr Val Leu Val Leu Val Leu Ala Val Cys Leu Leu Cys 260 265 270	
	Trp Thr Pro Phe His Leu Ala Ser Val Val Ala Leu Thr Thr Asp Leu 275 280 285	
15	Pro Gln Thr Pro Leu Val Ile Ser Met Ser Tyr Val Ile Thr Ser Leu 290 295 300	
	Thr Tyr Ala Asn Ser Cys Leu Asn Pro Phe Leu Tyr Ala Phe Leu Asp 305 310 315 320	
20	Asp Asn Phe Arg Lys Asn Phe Arg Ser Ile Leu Arg Cys 325 330	
25	(18) INFORMATION FOR SEQ ID NO:17: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 48 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17: ACGAATTCAG CCATGGTCCT TGAGGTGAGT GACCACCAAG TGCTAAAT	48
20		
30	(19) INFORMATION FOR SEQ ID NO:18: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 27 base pairs (B) TYPE: nucleic acid	
35	(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

WO 00/22129 PCT/US99/23938

15

GAGGATCCTG GAATGCGGGG AAGTCAG

5

27

(20) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1107 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

ATGGTCCTTG AGGTGAGTGA CCACCAAGTG CTAAATGACG CCGAGGTTGC CGCCCTCCTG 60 GAGAACTTCA GCTCTTCCTA TGACTATGGA GAAAACGAGA GTGACTCGTG CTGTACCTCC 120 CCGCCCTGCC CACAGGACTT CAGCCTGAAC TTCGACCGGG CCTTCCTGCC AGCCCTCTAC AGCCTCCTCT TTCTGCTGGG GCTGCTGGGC AACGGCGGG TGGCAGCCGT GCTGCTGAGC 240 CGGCGGACAG CCCTGAGCAG CACCGACACC TTCCTGCTCC ACCTAGCTGT AGCAGACACG 300 15 CTGCTGGTGC TGACACTGCC GCTCTGGGCA GTGGACGCTG CCGTCCAGTG GGTCTTTGGC 360 TCTGGCCTCT GCAAAGTGGC AGGTGCCCTC TTCAACATCA ACTTCTACGC AGGAGCCCTC 420 CTGCTGGCCT GCATCAGCTT TGACCGCTAC CTGAACATAG TTCATGCCAC CCAGCTCTAC 480 CGCCGGGGGC CCCCGGCCCG CGTGACCCTC ACCTGCCTGG CTGTCTGGGG GCTCTGCCTG 540 CTTTTCGCCC TCCCAGACTT CATCTTCCTG TCGGCCCACC ACGACGAGCG -CCTCAACGCC 20 ACCCACTGCC AATACAACTT CCCACAGGTG GGCCGCACGG CTCTGCGGGT GCTGCAGCTG 660 GTGGCTGGCT TTCTGCTGCC CCTGCTGGTC ATGGCCTACT GCTATGCCCA CATCCTGGCC GTGCTGCTGG TTTCCAGGGG CCAGCGGCGC CTGCGGGCCA TGCGGCTGGT GGTGGTGGTC 780 GTGGTGCCT TTGCCCTCTG CTGGACCCCC TATCACCTGG TGGTGCTGGT GGACATCCTC ATGGACCTGG GCGCTTTGGC CCGCAACTGT GGCCGAGAAA GCAGGGTAGA CGTGGCCAAG 900 25 TCGGTCACCT CAGGCCTGGG CTACATGCAC TGCTGCCTCA ACCCGCTGCT CTATGCCTTT 960 GTAGGGGTCA AGTTCCGGGA GCGGATGTGG ATGCTGCTCT TGCGCCTGGG CTGCCCCAAC 1020 CAGAGAGGC TCCAGAGGCA GCCATCGTCT TCCCGCCGGG ATTCATCCTG GTCTGAGACC 1080 TCAGAGGCCT CCTACTCGGG CTTGTGA 1107

(21) INFORMATION FOR SEQ ID NO:20:

5	(i)	(A) (B) (C)	LENCE TYPE STRA TOPO	STH: E: an ANDEI	368 mino ONESS	amin acio S:	no ao d	cids		. ,						
	(ii) MC	LECU	LE T	YPE:	pro	tein	• , '		٠.						
			QUEN													
	Met 1	Val	Leu		Val : 5	Ser .	Asp	His	Gln	Val :	Leu .	Asn .	Asp	Ala	Glu 15	Val
10	Ala	Ala	Leu	Leu 20	Glu .	Asn	Phe	Ser	Ser 25	Ser	Tyr	Asp	Tyr	Gly 30	Glu	Asn
	Glu	Ser	Asp 35	Ser	Cys	Cys	Thr	Ser 40	Pro	Pro	Cys	Pro	Gln 45	Asp	Phe	Ser
15	Leu	Asn 50	Phe	Asp	Arg	Ala	Phe 55	Leu	Pro	Ala	Leu	Tyr 60	Ser	Leu	Leu	Phe
	Leu 65	Leu	Gly	Leu	Leu	Gly 70	Asn	Gly	Ala	Val	Ala 75	Ala	Val	Leu	Leu	Ser 80
	Arg	Arg	Thr	Ala	Leu 85	Ser	Ser	Thr	Asp	Thr 90	Phe	Leu	Leu	His	Leu 95	Ala
20	Val	Ala	Asp	Thr 100	Leu	Leu	Val	Leu	Thr 105	Leu	Pro	Leu	Trp	Ala 110	Val	Asp
	Ala	Ala	Val	Gln	Trp	Val	Phe	Gly 120	Ser	Gly	Leu	Cys	Lys 125	Val	Ala	Gly
25	Ala	Lev	n Phe	Asn	Ile	Asn	Phe	Tyr	Ala	Gly	Ala	Leu 140	Leu	Leu	Ala	Cys
	Ile 14!		Phe	Asp	Arg	Tyr 150	Leu	Asn	Ile	Val	His 155	Ala	Thr	Gln	Leu	Tyr 160
	Arg	g Arg	g Gly	Pro	Pro		Arg	Val	Thr	Leu 170	Thr	Cys	Leu	ı Ala	Val 175	Trp
30	Gl	y Le	u Cys	Leu 180		Phe	. Ala	Leu	Pro 185	Asp	Phe	lle	Phe	190	Ser	Ala
	Hi	s Hi	s Asp 195		a Arg	, Lev	ı Ası	1 Ala 200	Thi	c His	Cys	; Glr	туі 20!	r Asr	n Phe	Pro
35	Gl	n Va 21	1 Gly 0	/ Arg	Thr	: Ala	1 Let 21!	ı Arg	y Va:	l Leu	ı Glr	Lei 220	ı Va:	l Ala	a Gly	Phe
	Le 22		u Pro	Let	ı Lev	ı Va:		t Ala	а Ту:	r Cys	235	r Ala	a Hi	s Ile	e Lev	1 Ala 240

DEIGNOCID AND DOCUMENTAL IA

WO 00/22129 PCT/US99/23938

	Val	Leu	Leu	Val	Ser 245	Arg	Gly	Gln	Àrg	Arg 250	Leu	Arg	Ala	Met	Arg 255	Leu	
	Val	Val	Val	Val 260	Val	Val	Ala	Phe	Ala 265	Leu	Cys	Trp	Thr	Pro 270	Tyr	His	
5	Leu	Val	Val 275	Leu	Val	Asp	Ile	Leu 280	Met	Asp	Leu	Gly	Ala 285	Leu	Ala	Arg	
	Asn	Cys 290	Gly	Arg	Glu	Ser	Arg 295	Val	Asp	Val	Ala	Lys 300	Ser	Val	Thr	Ser	
10	Gly 305		Gly	Tyr	Met	His 310	Cys	Cys	Leu	Asn	Pro 315	Leu	Leu	Tyr	Ala	Phe 320	
	Val	Gly	Val	Lys	Phe 325	Arg	Glu	Arg	Met	Trp 330	Met	Leu	Leu	Leu	Arg 335	Leu	
	Gly	Cys	Pro	Asn 340	Gln	Arg	Gly	Leu	Gln 345	Arg	Gln	Pro	Ser	Ser 350	Ser	Arg	
15	Arg	Asp	Ser 355		Ţŗp	Ser	Glu	Thr 360	Ser	Glu	Ala	Ser	Tyr 365	Ser	Gly	Leu	
•	(22) IN	FORM	ATIO	n fo	R SE	Q ID	NO:	21:									
20	(i	(A (B (C	QUEN) LE) TY) ST	NGTH PE : RAND	: 30 nucl EDNE	bas eic SS:	e pa acid sing	irs									
	(i		OLEC					enom	ic)								
			EQUE							NO:2	1:						
25	TTAAGCTI												•				30
	(23) INF	'ORMA	TION	FOR	SEQ	ID	NO : 2	2:									
30	(i	(A (E (C	QUEN LE L) LE L) TY L) SI L) TO	NGTH PE: RAND	: 30 nucl	bas eic SS:	e pa acid sing	irs									
	i)	i) M	OLEC	TULE	TYPE	: DN	IA (g	enom	ic)								
	()	ci) S	EQUE	NCE	DESC	RIPT	: NOI	SEC	ID.	NO:2	2:						
	TTGGATC	CAA A	AGAA	CCAT	G CA	CCTC	CAGAG	}									.30
35	(24) IN	FORMA	MOITA	FOF	SEC) ID	NO: 2	23 :									
	(i)	SEC	UENC	CE CH	LARAC	TERI	STIC	cs:									

- (A) LENGTH: 1074 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

	ATGGCTGATG	ACTATGGCTC	TGAATCCACA	TCTTCCATGG	AAGACTACGT	TAACTTCAAC	60
	TTCACTGACT	TCTACTGTGA	GAAAAACAAT	GTCAGGCAGT	TTGCGAGCCA	TTTCCTCCCA	120
	CCCTTGTACT	GGCTCGTGTT	CATCGTGGGT	GCCTTGGGCA	ACAGTCTTGT	TATCCTTGTC	180
10	TACTGGTACT	GCACAAGAGT	GAAGACCATG	ACCGACATGT	TCCTTTTGAA	TTTGGCAATT	240
	GCTGACCTCC	TCTTTCTTGT	CACTCTTCCC	TTCTGGGCCA	TTGCTGCTGC	TGACCAGTGG	300
	AAGTTCCAGA	CCTTCATGTG	CAAGGTGGTC	AACAGCATGT	ACAAGATGAA	CTTCTACAGC	360
,	TGTGTGTTGC	TGATCATGTG	CATCAGCGTG	GACAGGTACA	TTGCCATTGC	CCAGGCCATG	420
	AGAGCACATA	CTTGGAGGGA	GAAAAGGCTT	TTGTACAGCA	AAATGGTTTG	CTTTACCATC	480
15	TGGGTATTGG	CAGCTGCTCT	CTGCATCCCA	GAAATCTTAT	ACAGCCAAAT	CAAGGAGGAA	. 540
	TCCGGCATTG	CTATCTGCAC	CATGGTTTAC	CCTAGCGATG	AGAGCACCAA	ACTGAAGTCA	600
	GCTGTCTTGA	CCCTGAAGGT	CATTCTGGGG	TTCTTCCTTC	CCTTCGTGGT	CATGGCTTGC	660
	TGCTATACCA	TCATCATTCA	CACCCTGATA	CAAGCCAAGA	AGTCTTCCAA	GCACAAAGCC	. 720
	CTAAAAGTGA	CCATCACTGT	CCTGACCGTC	TTTGTCTTGT	CTCAGTTTCC	CTACAACTGC	780
20	ATTTTGTTGG	TGCAGACCAT	TGACGCCTAT	GCCATGTTCA	TCTCCAACTG	TGCCGTTTCC	840
	ACCAACATTG	ACATCTGCTT	CCAGGTCACC	CAGACCATCG	CCTTCTTCCA	CAGTTGCCTG	900
	AACCCTGTTC	TCTATGTTTT	TGTGGGTGAG	AGATTCCGCC	GGGATCTCGT	GAAAACCCTG	960
	AAGAACTTGG	GTTGCATCAG	CCAGGCCCAG	TGGGTTTCAT	TTACAAGGAG	AGAGGGAAGC	1020
	TTGAAGCTGT	CGTCTATGTT	GCTGGAGACA	ACCTCAGGAG	CACTCTCCCT	CTGA	1074

25 (25) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 357 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
- 30 (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein

	(x:	i) SI	EQUE	NCE I	DESCI	RIPT	ION:	SEQ	ID 1	NO : 24	1 :					
	Met 1	Ala	Asp	Asp	Tyr 5	Gly	Ser	Glu	Ser	Thr 10	Ser	Ser	Met	Glu	Asp 15	Tyr
5	Val	Asn	Phe	Asn 20	Phe	Thr	Asp	Phe	Tyr 25	Cys	Glu	Lys	Asn	Asn 30	Val	Arg
	Gln	Phe	Ala 35	Ser	His	Phe	Leu	Pro 40	Pro	Leu	Tyr	Trp	Leu 45	Val	Phe	Ile
	Val	Gly 50	Ala	Leu	Gly	Asn	Ser 55	Leu	Val	Ile	Leu	Val 60	Tyr	Trp	Tyr	Cys
10	Thr 65	Arg	Val	Lys	Thr	Met 70	Thr	Asp	Met	Phe	Leu 75	Leu	Asn	Leu	Ala	Ile 80
	Ala	Лsp	Leu	Leu	Phe 85	Leu	Val	Thr	Leu	Pro 90	Phe	Trp	Ala	Ile	Ala 95	Ala
15	Ala	Asp	Gln	Trp 100	Lys	Phe	Gln	Thr	Phe 105	Met	Cys	Lys	Val	Val 110	Asn	Ser
	Met	Tyr	Lys 115	Met	Asn	Phe	Tyr	Ser 120	Cys	Val	Leu	Leu	Ile 125	Met	Cys	Ile
	Ser	Val 130	Asp	Arg	Tyr	Ile	Ala 135		Ala	Gln	Ala	Met 140	Arg	Ala	His	Thr
20	Trp 145	Arg	Glu	Lys	Arg	Leu 150	Leu	Tyr	Ser	Lys	Met 155	Val	Cys	Phe	Thr	Ile 160
	Trp	Val	Leu	Ala	Ala 165	Ala	Leu	Cys	Ile	Pro 170	Glu	Ile	Leu	Tyr	Ser 175	Gln
25	Ile	Lys	Glu	Glu 180	Ser	Gly	Ile	Ala	Ile 185	Cys	Thr	Met	Val	Tyr 190	Pro	Ser
	Asp	Glu	Ser 195	Thr	Lys	Leu	Lys	Ser 200	Ala	Val	Leu	Thr	Leu 205	Lys	Val	Ile
	Leu	Gly 210	Phe	Phe	Leu	Pro	Phe 215	Val	Val	Met	Ala	Cys 220	Cys	Tyr	Thr	Ile
30	Ile 225	Ile	His	Thr	Leu	Ile 230	Gln	Ala	Lys	Lys	Ser 235	Ser	Lys	His	Lys	Ala 240
	Leu	Lys	Val	Thr	Ile 245	Thr	Val	Leu	Thr	Val 250	Phe	Val	Leu	Ser	Gln 255	Phe
35	Pro	Tyr	Asn	Cys 260	Ile	Leu	Leu	Val	Gln 265	Thr	Ile	Asp	Ala	Tyr 270	Ala	Met
	Phe	Ile	Ser 275	Asn	Cys	Ala	Val	Ser 280	Thr	Asn	Ile	Asp	Ile 285	Cys	Phe	Gln

	20
	Val Thr Gln Thr Ile Ala Phe Phe His Ser Cys Leu Asn Pro Val Leu 290 295 300
	Tyr Val Phe Val Gly Glu Arg Phe Arg Arg Asp Leu Val Lys Thr Leu 305 310 315 320
5	Lys Asn Leu Gly Cys Ile Ser Gln Ala Gln Trp Val Ser Phe Thr Arg 325 330 335
	Arg Glu Gly Ser Leu Lys Leu Ser Ser Met Leu Leu Glu Thr Thr Ser 340 345 350
10	Gly Ala Leu Ser Leu 355
	(26) INFORMATION FOR SEQ ID NO:25:
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1110 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear
	(ii) MOLECULE TYPE: DNA (genomic)
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:
	ATGGCCTCAT CGACCACTCG GGGCCCCAGG GTTTCTGACT TATTTTCTGG GCTGCCGCCG 60
20	GCGGTCACAA CTCCCGCCAA CCAGAGCGCA GAGGCCTCGG CGGGCAACGG GTCGGTGGCT 120
	GGCGCGGACG CTCCAGCCGT CACGCCCTTC CAGAGCCTGC AGCTGGTGCA TCAGCTGAAG 180
	GGGCTGATCG TGCTGCTCTA CAGCGTCGTG GTGGTCGTGG GGCTGGTGGG CAACTGCCTG 240
	CTGGTGCTGG TGATCGCGCG GGTGCCGCGG CTGCACAACG TGACGAACTT CCTCATCGGC 300
	AACCTGGCCT TGTCCGACGT GCTCATGTGC ACCGCCTGCG TGCCGCTCAC GCTGGCCTAT 360
25	GCCTTCGAGC CACGCGGCTG GGTGTTCGGC GGCGGCCTGT GCCACCTGGT CTTCTTCCTG 420
	CAGCCGGTCA CCGTCTATGT GTCGGTGTTC ACGCTCACCA CCATCGCAGT GGACCGCTAC 480

ACCCAGAGCC AGGCCGACTG GGACCGCGCT CGGCGCCGGC GCACCTTCTG CTTGCTGGTG 840

GTCGTGCTGG TGCACCCGCT GAGGCGCGCA TCTCGCTGCG CCTCAGCCTA CGCTGTGCTG

GCCATCTGGG CGCTGTCCGC GGTGCTGGCG CTGCCGCCCG CCGTGCACAC CTATCACGTG

GAGCTCAAGC CGCACGACGT GCGCCTCTGC GAGGAGTTCT GGGGCTCCCA GGAGCGCCAG

CTCCTGTCTT ACGTCCGGGT GTCAGTGAAG CTCCGCAACC GCGTGGTGCC GGGCTGCGTG

30 CGCCAGCTCT ACGCCTGGGG GCTGCTGCTG GTCACCTACC TGCTCCCTCT GCTGGTCATC

540

600

660

720

WO 00/22129 PCT/US99/23938

21

	GTGGTCGT	GG TO	GTG:	rtcg	C CG1	rctg	CTGG	CTG	CCGC	rgc 1	ACGT	CTTC	AA C	CTGC	rgcgo	3 9	900
	GACCTCGAC	cc co	CAC	GCCA?	r cga	ACCC!	TTAC	GCC	rttgo	GC :	rggto	CAG	CT G	CTCTC	CCA	2 5	960
	TGGCTCGC	CA TO	GAGT	rcgg	CTC	CTA	CAAC	CCC	TCAT	CT 1	ACGC	CTGG	CT G	CACGI	ACAGO	10	020
	TTCCGCGA	GG AC	CTG	CGCA	A· ACI	rgtt	GTC	GCT	rggco	ccc o	CAAC	GATA	SC C	CCCZ	ATGG	2 10	080
5	CAGAATAT	GA CO	CGTC	AGCG:	r ggr	CAT	CTGA									13	L10
	(27) INFO	ORMA!	rion	FOR	SEQ	ID 1	NO : 26	5:									
10	(i)	(B)	QUENC LEI TYI STI	NGTH PE: & RANDI	: 369 amino EDNES	e am: cac: SS:	ino a id	acids	5								٠
	(i:	i) MO	OLECT	JLE :	TYPE:	pro	oteir	n									
	(x:	i) SI	EQUE	NCE I	DESC	RIPT	ION:	SEQ	ID 1	10:26	5:						-
15	Met 1	Ala	Ser	Ser	Thr 5	Thr	Arg	Gly	Pro	Arg 10	Val	Ser	Asp	Leu	Phe 15	Ser	
	Gly	Leu	Pro	Pro 20	Ala	Val	Thr	Thr	Pro 25	Ala	Asn	Gln	Ser	Ala 30	Glu	Ala	, S.
	Ser	Ala	Gly 35	Asn	Gly	Ser	Val	Ala 40	Gly	Ala	Asp	Ala	Pro 45	Ala	Val	Thr	
20	Pro	Phe 50	Gln	Ser	Leu	Gln	Leu 55	Val	His	Gln	Leu	Lys 60	Gly	Leu	Ile	Val	
	Leu 65	Leu	Tyr	Ser	Val	Val 70	Val	Val	Val	Gly	Leu 75	Val	Gly	Asn	Cys	Leu 80	
25	Leu	Val	Leu	Val	Ile 85	Ala	Arg	Val	Pro	Arg 90	Leu	His	Asn	Val	Thr 95	Asn	
	Phe	Leu	Ile	Gly 100	Asn	Leu	Ala	Leu	Ser 105	Asp	Val	Leu	Met	Cys 110	Thr	Ala	1
	Cys	Val	Pro 115	Leu	Thr	Leu	Ala	Tyr 120	Ala	Phe	Glu	Pro	Arg 125	Gly	Trp	Val	
30	Phe	Gly 130	Gly	Gly	Leu		His 135	Leu	Val	Phe	Phe	Leu 140	Gln	Pro	Val	Thr	
	Val 145	Tyr	Val	Ser	Val	Phe 150	Thr	Leu	Thr	Thr	Ile 155	Ala	Val	Asp	Arg	Tyr 160	
35	Val	Val	Leu	Val	His 165	Pro	Leu	Arg	Arg	Ala 170	Ser	Arg	Cys	Ala	Ser 175	Ala	

		Tyr	Ala	Val	Leu 180	Ala	Ile	Trp	Ala	Leu 185	Ser	Ala	Val	Leu	Ala 190	Leu	Pro	
		Pro	Ala	Val 195	His	Thr	Tyr	His	Val 200	Glu	Leu	Lys	Pro	His 205	Asp	Val	Arg	
5		Leu	Cys 210	Glu	Glu	Phe	Trp	Gly 215	Ser	Gln	Glu	Arg.	Gln 220	Arg	Gln	Leu	Tyr	
		Ala 225	Trp	Gly	Leu	Leu	Leu 230	Val	Thr	Tyr	Leu	Leu 235	Pro	Leu	Leu	Val	Ile 240	
10		Leu	Leu	Ser	Tyr	Val 245	Arg	Val	Ser	Val	Lys 250	Leu	Arg	Asn	Arg	Val 255	Val	
		Pro	Gly	Cys	Val 260	Thr	Gln	Ser	Gln	Ala 265	Asp	Trp	Asp	Arg	Ala 270	Arg	Arg	
		Arg	Arg	Thr 275		Cys	Leu	Leu	Val 280	Val	Val	Val	Val	Val 285	Phe	Ala	Val	
15		Cys	Trp 290	Leu	Pro	Leu	His	Val 295		Asn	Leu	Leu	Arg 300	Asp	Leu	Asp	Pro	
		His 305	Ala	Ile	Asp	Pro	Tyr 310		Phe	Gly	Leu	Val 315	Gln	Leu	Leu	Cys	His 320	
20	•	Trp	Leu	Ala	Met	Ser 325		Ala	. Cys	Tyr	Asn 330		Phe	Ile	Tyr	Ala 335	Trp	
		Leu	His	Asp	Ser 340		Arg	Glu	Glu	1 Leu 345		Lys	Leu	Leu	Val 350	Ala	Trp	
		Pro	Arg	Lys 355		Ala	Pro	His	Gly 360		Asn	Met	Thr	Val 365	Ser	Val	Val	
25		Ile	e															
	(28)	IN	FORM	TION	1 FOF	SEÇ	Q ID	NO:2	27:									
30		(:	(E	A) Li 3) T C) S'	NCE C ENGTH YPE: TRANI OPOLO	i: 10 nuc: DEDNI	083 l Leic ESS:	acio sing	pai: d	rs								
	٠.	(ii) P	MOLE	CULE	TYP	E: D	NA (geno	mic)								
		. (:	xi) :	SEQU	ENCE	DES	CRIP	TION	: SE	Q ID	NO:	27:						
35	ATG	BACC	CAG 2	AAGA	AACT'	TC A	GTTT	ATTT	G GA	TTAT	TACT	ATG	CTAC	GAG (CCCA	AACT	CT	60
	GAC	ATCA	GGG :	AGAC	CCAC	TC C	CATG	TTCC	T TA	CACC	TCTG	TCT	TCCT	TCC 2	AGTC'	TTTT.	AC	120

PCT/US99/23938

	ACAGCTGTGT	TCCTGACTGG	AGTGCTGGGG	AACCTTGTTC	TCATGGGAGC	GTTGCATTTC	180
	AAACCCGGCA	GCCGAAGACT	GATCGACATC	TTTATCATCA	ATCTGGCTGC	CTCTGACTTC	240
	ATTTTTCTTG	TCACATTGCC	TCTCTGGGTG	GATAAAGAAG	CATCTCTAGG	ACTGTGGAGG	300
	ACGGGCTCCT	TCCTGTGCAA	AGGGAGCTCC	TACATGATCT	CCGTCAATAT	GCACTGCAGT	360
5	GTCCTCCTGC	TCACTTGCAT	GAGTGTTGAC	CGCTACCTGG	CCATTGTGTG	GCCAGTCGTA	420
	TCCAGGAAAT	TCAGAAGGAC	AGACTGTGCA	TATGTAGTCT	GTGCCAGCAT	CTGGTTTATC	480
	TCCTGCCTGC	TGGGGTTGCC	TACTCTTCTG	TCCAGGGAGC	TCACGCTGAT	TGATGATAAG	540
	CCATACTGTG	CAGAGAAAAA	GGCAACTCCA	ATTAAACTCA	TATGGTCCCT	GGTGGCCTTA	600
	ATTTTCACCT	TTTTTGTCCC	TTTGTTGAGC	ATTGTGACCT	GCTACTGTTG	CATTGCAAGG	660
0	AAGCTGTGTG	CCCATTACCA	GCAATCAGGA	AAGCACAACA	AAAAGCTGAA	GAAATCTATA	720
	AAGATCATCT	TTATTGTCGT	GGCAGCCTTT	CTTGTCTCCT	GGCTGCCCTT	CAATACTTTC	780
	AAGTTCCTGG	CCATTGTCTC	TGGGTTGCGG	CAAGAACACT	ATTTACCCTC	AGCTATTCTT	840
	CAGCTTGGTA	TGGAGGTGAG	TGGACCCTTG	GCATTTGCCA	ACAGCTGTGT	CAACCCTTTC	/ 900
	ATTTACTATA	TCTTCGACAG	CTACATCCGC	CGGGCCATTG	TCCACTGCTT	GTGCCCTTGC	960
5	CTGAAAAACT	ATGACTTTGG	GAGTAGCACT	GAGACATCAG	ATAGTCACCT	CACTAAGGCT	1020
	CTCTCCACCT	TCATTCATGC	AGAAGATTTT	GCCAGGAGGA	GGAAGAGGTC	TGTGTCACTC	1080
	TAA			. • . ,			1083
	(29) INFOR	MATION FOR	SEQ ID NO:2	8:			
20	(i)	SEQUENCE CH (A) LENGTH: (B) TYPE: a	360 amino				

- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Met Asp Pro Glu Glu Thr Ser Val Tyr Leu Asp Tyr Tyr Tyr Ala Thr 1 5 10 15

Ser Pro Asn Ser Asp Ile Arg Glu Thr His Ser His Val Pro Tyr Thr 20 25 30

30 Ser Val Phe Leu Pro Val Phe Tyr Thr Ala Val Phe Leu Thr Gly Val 35 40 45

	Leu	Gly 50	Asn	Leu	Val	Leu	Met 55	Gly	Ala	Leu	His	Phe 60	Lys	Pro	Gly	Ser
	Arg 65	Arg	Leu	Ile	Asp	Ile 70	Phe	Ile	Ile	Asn	Leu 75	Ala	Ala	Ser	Asp	Phe 80
5	Ile	Phe	Leu	Val	Thr 85	Leu	Pro	Leu	Trp	Val 90	Asp	Lys	Glu	Ala	Ser 95	Leu
	Gly	Leu	Trp	Arg 100	Thr	Gly	Ser	Phe	Leu 105	Cys	Lys	Gly	Ser	Ser 110	Tyr	Met
10	Ile	Ser	Val 115	Asn	Met	His	Cys	Ser 120	Val	Leu	Leu	Leu	Thr 125	Cys	Met	Ser
	Val	Asp 130	Arg	Tyr	Leu		Ile 135	Val	Trp	Pro	Val	Val 140	Ser	Arg	Lys	Phe
	Arg 145	Arg	Thr	Asp	Cys	Ala 150	Tyr	Val	Val	Cys	Ala 155	Ser	Ile	Trp	Phe	Ile 160
15	Ser	Cys	Leu	Leu	Gly 165	Leu	Pro	Thr	Leu	Leu 170	Ser	Arg	Glu	Leu	Thr 175	Leu
	Ile	Asp	Asp	Lys 180	Pro	Tyr	Cys	Ala	Glu 185	Lys	Lys	Ala	Thr	Pro 190	Ile	Lys
20	Leu	Ile	Trp 195	Ser	Leu	Val	Ala	Leu 200	Ile	Phe	Thr	Phe	Phe 205	Val	Pro	Leu
		210					215					Arg 220				
	His 225	Tyr	Gln	Gln	Ser	Gly 230	Lys	His	Asn	Lys	Lys 235	Leu	Lys	Lys	Ser	Ile 240
	Lys	Ile	Ile	Phe	Ile 245	Val	Val	Ala	Ala	Phe 250	Leu	Val	Ser	Trp	Leu 255	Pro
	Phe	Asn	Thr	Phe 260	Lys	Phe	Leu	Ala	Ile 265	Val	Ser	Gly	Leu	Arg 270	Gln	Glu
30		-	275					280				Met	285			
		290				•	295					Phe 300				
	305	<u> </u>		-		310		•			315	Cys				320
35	•	-			325					330		Thr			335	
	Leu	Thr	Lys	Ala	ren	ser	Thr	Pne	тTе	Hls	Ala	Glu	Asp	rne	Ата	Arg

WO 00/22129 PCT/US99/23938

	340 345	350	
	Arg Arg Lys Arg Ser Val Ser Leu 355 360		
	(30) INFORMATION FOR SEQ ID NO:29:		
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 		
10	(ii) MOLECULE TYPE: DNA (genomic)		
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:		
	CTAGAATTCT GACTCCAGCC AAAGCATGAA T		31
	(31) INFORMATION FOR SEQ ID NO:30:		
15	(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 30 base pairs(B) TYPE: nucleic acid		÷.
	(C) STRANDEDNESS: single (D) TOPOLOGY: linear	•	.
	(ii) MOLECULE TYPE: DNA (genomic)		
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:		
	GCTGGATCCT AAACAGTCTG CGCTCGGCCT		30
	(32) INFORMATION FOR SEQ ID NO:31:		
25	(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 1020 base pairs(B) TYPE: nucleic acid		
20	(C) STRANDEDNESS: single (D) TOPOLOGY: linear		
	(ii) MOLECULE TYPE: DNA (genomic)		
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:		
30	ATGAATGGCC TTGAAGTGGC TCCCCCAGGT CTGATCACCA ACTTCTCCCT	GGCCACGGCA	60
	GAGCAATGTG GCCAGGAGAC GCCACTGGAG AACATGCTGT TCGCCTCCTT	CTACCTTCTG	120
	GATTTTATCC TGGCTTTAGT TGGCAATACC CTGGCTCTGT GGCTTTTCAT	CCGAGACCAC	180
	AAGTCCGGGA CCCCGGCCAA CGTGTTCCTG ATGCATCTGG CCGTGGCCGA	CTTGTCGTGC	240
	GTGCTGGTCC TGCCCACCCG CCTGGTCTAC CACTTCTCTG GGAACCACTG	GCCATTTGGG	300
	· · · · · · · · · · · · · · · · · · ·		

	GAAATCGCAT	GCCGTCTCAC	CGGCTTCCTC	TTCTACCTCA	ACATGTACGC	CAGCATCTAC	360
	TTCCTCACCT	GCATCAGCGC	CGACCGTTTC	CTGGCCATTG	TGCACCCGGT	CAAGTCCCTC	420
	AAGCTCCGCA	GGCCCCTCTA	CGCACACCTG	GCCTGTGCCT	TCCTGTGGGT	GGTGGTGGCT	480
	GTGGCCATGG	CCCCGCTGCT	GGTGAGCCCA	CAGACCGTGC	AGACCAACCA	CACGGTGGTC	540
5	TGCCTGCAGC	TGTACCGGGA	GAAGGCCTCC	CACCATGCCC	TGGTGTCCCT	GGCAGTGGCC	600
	TTCACCTTCC	CGTTCATCAC	CACGGTCACC	TGCTACCTGC	TGATCATCCG	CAGCCTGCGG	660
	CAGGGCCTGC	GTGTGGAGAA	GCGCCTCAAG	ACCAAGGCAG	TGCGCATGAT	CGCCATAGTG	720
	CTGGCCATCT	TCCTGGTCTG	CTTCGTGCCC	TACCACGTCA	ACCGCTCCGT	CTACGTGCTG	780
	CACTACCGCA	GCCATGGGGC	CTCCTGCGCC	ACCCAGCGCA	TCCTGGCCCT	GGCAAACCGC	840
10	ATCACCTCCT	GCCTCACCAG	CCTCAACGGG	GCACTCGACC	CCATCATGTA	TTTCTTCGTG	.900
	GCTGAGAAGT	TCCGCCACGC	CCTGTGCAAC	TTGCTCTGTG	GCAAAAGGCT	CAAGGGCCCG	960
	CCCCCAGCT	TCGAAGGGAA	AACCAACGAG	AGCTCGCTGA	GTGCCAAGTC	AGAGCTGTGA	1020
				•			

(33) INFORMATION FOR SEQ ID NO:32:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 339 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Met Asn Gly Leu Glu Val Ala Pro Pro Gly Leu Ile Thr Asn Phe Ser

1 10 15

Leu Ala Thr Ala Glu Gln Cys Gly Gln Glu Thr Pro Leu Glu Asn Met 20 25 30

25 Leu Phe Ala Ser Phe Tyr Leu Leu Asp Phe Ile Leu Ala Leu Val Gly 35 40 45

Asn Thr Leu Ala Leu Trp Leu Phe Ile Arg Asp His Lys Ser Gly Thr 50 55 60

Pro Ala Asn Val Phe Leu Met His Leu Ala Val Ala Asp Leu Ser Cys 65 70 75 80

Val Leu Val Leu Pro Thr Arg Leu Val Tyr His Phe Ser Gly Asn His 85 90 95

Trp Pro Phe Gly Glu Ile Ala Cys Arg Leu Thr Gly Phe Leu Phe Tyr

27

					100					.105	•		-		110		
		Leu	Asn	Met 115	Tyr	Ala	Ser	Ile	Туг 120	Phe	Leu	Thr	Cys	Ile 125	Ser	Ala	Asp
5		Arg	Phe 130	Leu	Ala	Ile	Val	His 135	Pro	Val	Lys	Ser	Leu 140	Lys	Leu	Arg	Arg
		Pro 145	Leu	Tyr	Ala	His	Leu 150	Ala	Cys	Ala	Phe	Leu 155	Trp	Val	Val	Val	Ala 160
		Val	Ala	Met	Ala	Pro 165	Leu	Leu	Val	Ser	Pro 170	Gln	Thr	Val	Gln	Thr 175	Asn
10		His	Thr	Val	Val 180	Cys	Leu	Gln	Leu	Tyr 185	Arg	Glu	Lys	Ala	Ser 190	His	His
		Ala	Leu	Val 195	Ser	Leu	Ala	Val	Ala 200	Phe	Thr	Phe	Pro	Phe 205	Ile	Thr	Thr
15		Val	Thr 210	Cys	Tyr	Leu	Leu	Ile 215	Ile	Arg	Ser	Leu	Arg 220	Gln	Gly	Leu	Arg
		Val 225	Glu	Lys	Arg	Leu	Lys 230	Thr	Lys	Ala	Val	Arg 235	Met	Ile	Ala	Ile	Val 240
		Leu	Ala	Ile	Phe	Leu 245	Val	Cys	Phe	Val	Pro 250	Tyr	His	Val	Asn	Arg 255	Ser
20		Val	Tyr	Val	Leu 260	His	Týr	Arg	Ser	His 265	Gly	Ala	Ser	Cys	Ala 270	Thr	Gln
		Arg	Ile	Leu 275	Ala	Leu	Ala	Asn	Arg 280	Ile	Thr	Ser	Cys	Leu 285	Thr	Ser	Leu
25		Asn	Gly 290	Ala	Leu	Asp	Pro	Ile 295	Met	Tyr	Phe	Phe	Val 300	Ala	Glu	Lys	Phe
		Arg 305	His	Ala	Leu	Cys	Asn 310	Leu	Leu	Cys	Gly	Lys 315	Arg	Leu	Lys	Gly	Pro 320
		Pro	Pro	Ser	Phe	Glu 325	Gly	Lys	Thr	Asn	Glu 330	Ser	Ser	Leu	Ser	Ala 335	Lys
30		Ser	Glu	Leu													
	(34)	INFO	RMAT	rion	FOR	SEQ	ID 1	10:33	3:								
35		(i)	(A) (B) (C)	UENC LEN TYP STF TOP	IGTH: PE: r RANDE	29 nucle	base eic a SS: s	e pai acid singl	irs								
			,														

(ii) MOLECULE TYPE: DNA (genomic)

	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:	
	ATAAGATGAT CACCCTGAAC AATCAAGAT	29
	(35) INFORMATION FOR SEQ ID NO:34:	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
10	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:	
	TCCGAATTCA TAACATTTCA CTGTTTATAT TGC	33
	(36) INFORMATION FOR SEQ ID NO:35:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 996 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:	
20	ATGATCACCC TGAACAATCA AGATCAACCT GTCACTTTTA ACAGCTCACA TCCAGATGAA	60
	TACAAAATTG CAGCCCTTGT CTTCTATAGC TGTATCTTCA TAATTGGATT ATTTGTTAAC	120
	ATCACTGCAT TATGGGTTTT CAGTTGTACC ACCAAGAAGA GAACCACGGT AACCATCTAT	180
	ATGATGAATG TGGCATTAGT GGACTTGATA TTTATAATGA CTTTACCCTT TCGAATGTTT	240
	TATTATGCAA AAGATGCATG GCCATTTGGA GAGTACTTCT GCCAGATTAT TGGAGCTCTC	300
25	ACAGTGTTTT ACCCAAGCAT TGCTTTATGG CTTCTTGCCT TTATTAGTGC TGACAGATAC	360
	ATGGCCATTG TACAGCCGAA GTACGCCAAA GAACTTAAAA ACACGTGCAA AGCCGTGCTG	420
	GCGTGTGTGG GAGTCTGGAT AATGACCCTG ACCACGACCA CCCCTCTGCT ACTGCTCTAT	480
	AAAGACCCAG ATAAAGACTC CACTCCCGCC ACCTGCCTCA AGATTTCTGA CATCATCTAT	540
	CTAAAAGCTG TGAACGTGCT GAACCTCACT CGACTGACAT TTTTTTTCTT GATTCCTTTG	600
30	TTCATCATGA TTGGGTGCTA CTTGGTCATT ATTCATAATC TCCTTCACGG CAGGACGTCT	660
	AAGCTGAAAC CCAAAGTCAA GGAGAAGTCC ATAAGGATCA TCATCACGCT GCTGGTGCAG	720

באובטטטוט אויט מיטטזטניאן ואי

GTGC	TCGT	'CT G	CTTI	ATGC	C CI	TCCA	CATC	TGT	TTC	CTT	TCCI	GATO	CT C	GGAA	CGGG	:G	780
GAGA	ACAG	TT A	CAAT	CCCI	'G GG	GAGC	CTTT	ACC	'ACC'I	TCC	TCAT	GAAC	CT C	CAGCA	CGTG	T	840
CTGG	ATGT	GA I	TCTC	TACT	'A CA	TCGT	TTCA	AAA	CAAT	TTC	AGGC	TCGA	GT C	CATTA	GTGT	'C	900
ATGC	TATA	.CC G	TAAT	TACC	T TC	GAAG	CCTG	CGC	'AGAA	AAA .	GTTT	'CCGA	TC I	GGTA	GTCT	'A	960
AGGT	CACT	AA G	CAAT	'ATAA	A CA	GTGA	AATG	TTA	TGA								996
(37)	INF	ORMA	TION	FOR	SEQ	ID	NO:3	6:									
		(A (B (C (D) LE) TY) ST) TO	CE C NGTH PE: RAND POLO ULE	: 33 amin EDNE GY:	1 am o ac SS: not	ino id rele	acid vant									
						_					_						
				NCE :				-									
	Met 1	Ile	Thr	Leu	Asn 5	Asn	Gln	Asp	Gln	Pro 10	Val	Thr	Phe	Asn	Ser 15	Ser	
	His	Pro	Asp	Glu 20	Tyr	Lys	Ile	Ala	Ala 25	Leu	Val	Phe	Tyr	Ser 30	Cys	Ile	
	Phe	Ile	Ile 35	Gly	Leu	Phe	Val	Asn 40	Ile	Thr	Ala	Leu	Trp 45	Val	Phe	Ser	
	Cys	Thr 50	Thr	Lys	Lys	Arg	Thr 55	Thr	Val	Thr	Ile	Tyr 60	Met	Met	Asn	Val	• .
	Ala 65	Leu	Val	Asp	Leu	Ile 70	Phe	Ile	Met	Thr	Leu 75	Pro	Phe	Arg	Met	Phe 80	
	Tyr	Tyr	Ala	Lys	Asp 85	Ala	Trp	Pro	Phe	Gly 90	Glu	Tyr	Phe	Cys	Gln 95	Ile	
	Ile	Gly	Ala	Leu 100	Thr	Val	Phe	Tyr	Pro 105	Ser	Ile	Ala	Leu	Trp 110	Leu	Leu	
	Ala	Phe	Ile 115	Ser	Ala	Asp	Arg	Туг 120	Met	Ala	Ile	Val	Gln 125	Pro	Lys	Tyr	
•	Ala	Lys 130	Glu	Leu	Lys	Asn	Thr 135	Cys	Lys	Ala	Val	Leu 140	Ala	Cys	Val	Gly	
	Val 145	Trp	Ile	Met	Thr	Leu 150	Thr	Thr	Thr	Thr	Pro 155	Leu	Leu	Leu	Leu	Tyr 160	
	Lys	Asp	Pro	Asp	Lys 165	Asp	Ser	Thr	Pro	Ala 170	Thr	Cys	Leu	Lys	Ile 175	Ser	

		Asp	Ile	Ile	Tyr 180	Leu	Lys	Ala	Val	Asn 185	Val	Leu	Asn	Leu	Thr 190	Arg	Leu
		Thr	Phe	Phe 195	Phe	Leu	Ile	Pro	Leu 200	Phe	Ile	Met	Ile	Gly 205	Cys	Tyr	Leu
5		Val	Ile 210	Ile	His	Asn	Leu	Leu 215	His	Gly	Arg	Thr	Ser 220	Lys	Leu	Lys	Pro
		Lys 225	Val	Lys	Glu	Lys	Ser 230	Ile	Arg	Ile	Ile	Ile 235	Thr	Leu	Leu	Val	Gln 240
10		Val	Leu	Val	Суѕ	Phe 245	Met	Pro	Phe	His	Ile 250	Cys	Phe	Ala	Phe	Leu 255	Met
		Leu	Gly	Thr	Gly 260	Glu	Asn	Ser	Tyr	Asn 265	Pro	Trp	Gly	Ala	Phe 270	Thr	Thr
		Phe	Leu	Met 275	Asn	Leu	Ser	Thr	Cys 280	Leu	Asp	Val	Ile	Leu 285	Tyr	Tyr	Ile
15		Val	Ser 290	Lys	Gln	Phe	Gln	Ala 295	Arg	Val	Ile	Ser	Val 300	Met	Leu	.Tyr	Arg
,		Asn 305	Tyr	Leu	Arg	Ser	Leu 310	Arg	Arg	Lys	Ser	Phe 315	Arg	Ser	Gly	Ser	Leu 320
20		Arg	Ser	Leu	Ser	Asn 325	Ile	Asn	Ser	Glu	Met 330	Leu			٠		
	(38)	INF	ORMA:	rion	FOR	SEQ	ID 1	10:3	7:								
25		(i)	(A) (B) (C)	LEI TYI	CE CH NGTH: PE: 1 RANDI POLOC	: 28 nucle EDNES	base eic a SS: s	e pa: acid sing:	irs								
		(i:	i) M(OLECT	JLE 1	YPE:	: DNA	A (ge	enom:	ic)							
		(x:	i) SI	EQUEI	NCE I	DESCI	RIPT	ON:	SEQ	ID 1	VO:37	7:					
	CCAA	GCTT(CC AC	GCC'	rggg	G TG:	rgcT(€G			4						28
30	(39)	INF	ORMA!	rion	FOR	SEQ	ID 1	10:31	3:								
35		(i)	(A) (B) (C)	LEI TYI	CE CH NGTH: PE: 1 RANDI POLOC	: 29 nucle EDNES	base eic a SS: s	e pa: acid sing:	irs				,			·	
		(i:	i) M	OLECT	JLE 1	TYPE	: DNI	4 (g	enom:	ic)							
		1	: \	יפוזסים	יורים ז	יביפרי	ייים דכ	CONT -	CEC.	י חד	ın. 20	٥.					

PCT/US99/23938

720

780

840

960

1077

5

ATGGATCCTG ACCTTCGGCC CCTGGCAGA 29 (40) INFORMATION FOR SEQ ID NO:39: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1077 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (genomic) (xi) SEQUENCE DESCRIPTION: SEQ ID NO:39: ATGCCCTCTG TGTCTCCAGC GGGGCCCTCG GCCGGGGCAG TCCCCAATGC CACCGCAGTG 60 ACAACAGTGC GGACCAATGC CAGCGGGCTG GAGGTGCCCC TGTTCCACCT GTTTGCCCGG 120 CTGGACGAGG AGCTGCATGG CACCTTCCCA GGCCTGTGCG TGGCGCTGAT GGCGGTGCAC 180 GGAGCCATCT TCCTGGCAGG GCTGGTGCTC AACGGGCTGG CGCTGTACGT CTTCTGCTGC 240 CGCACCCGGG CCAAGACACC CTCAGTCATC TACACCATCA ACCTGGTGGT GACCGATCTA 300 CTGGTAGGGC TGTCCCTGCC CACGCGCTTC GCTGTGTACT ACGGCGCCAG GGGCTGCCTG 360 CGCTGTGCCT TCCCGCACGT CCTCGGTTAC TTCCTCAACA TGCACTGCTC CATCCTCTTC 420 CTCACCTGCA TCTGCGTGGA CCGCTACCTG GCCATCGTGC GGCCCGAAGG CTCCCGCCGC 480 TGCCGCCAGC CTGCCTGTGC CAGGGCCGTG TGCGCCTTCG TGTGGCTGGC CGCCGGTGCC 540 GTCACCCTGT CGGTGCTGGG CGTGACAGGC AGCCGGCCCT GCTGCCGTGT CTTTGCGCTG 600 ACTGTCCTGG AGTTCCTGCT GCCCCTGCTG GTCATCAGCG TGTTTACCGG CCGCATCATG 660

GTGGCCGTGA CCCTCAGCAG CCTCAACAGC TGCATGGACC CCATCGTCTA CTGCTTCGTC 900 ACCAGTGGCT TCCAGGCCAC CGTCCGAGGC CTCTTCGGCC AGCACGGAGA GCGTGAGCCC

TGTGCACTGT CGCGGCCGGG TCTGCTCCAC CAGGGTCGCC AGCGCCGCGT GCGGGCCATG

CAGCTCCTGC TCACGGTGCT CATCATCTTT CTCGTCTGCT TCACGCCCTT CCACGCCCGC

CTCAGTGCCG GCCTCACCC CCTCACCCAG GCCTGGCTA ATGGGCCCGA GGCTTAG

AGCAGCGGTG ACGTGGTCAG CATGCACAGG AGCTCCAAGG GCTCAGGCCG TCATCACATC 1020

(41) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 358 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: (D) TOPOLOGY: not relevant (ii) MOLECULE TYPE: protein 5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:40: Met Pro Ser Val Ser Pro Ala Gly Pro Ser Ala Gly Ala Val Pro Asn 5 Ala Thr Ala Val Thr Thr Val Arg Thr Asn Ala Ser Gly Leu Glu Val 25 10 Pro Leu Phe His Leu Phe Ala Arg Leu Asp Glu Glu Leu His Gly Thr 40 Phe Pro Gly Leu Cys Val Ala Leu Met Ala Val His Gly Ala Ile Phe 55 Leu Ala Gly Leu Val Leu Asn Gly Leu Ala Leu Tyr Val Phe Cys Cys 15 75 Arg Thr Arg Ala Lys Thr Pro Ser Val Ile Tyr Thr Ile Asn Leu Val 90 Val Thr Asp Leu Leu Val Gly Leu Ser Leu Pro Thr Arg Phe Ala Val 105 20 Tyr Tyr Gly Ala Arg Gly Cys Leu Arg Cys Ala Phe Pro His Val Leu Gly Tyr Phe Leu Asn Met His Cys Ser Ile Leu Phe Leu Thr Cys Ile 135 140 . Cys Val Asp Arg Tyr Leu Ala Ile Val Arg Pro Glu Ala Pro Ala Ala 25 150 155 Cys Arg Gln Pro Ala Cys Ala Arg Ala Val Cys Ala Phe Val Trp Leu 165 Ala Ala Gly Ala Val Thr Leu Ser Val Leu Gly Val Thr Gly Ser Arg 185 30 Pro Cys Cys Arg Val Phe Ala Leu Thr Val Leu Glu Phe Leu Leu Pro 200 Leu Leu Val Ile Ser Val Phe Thr Gly Arg Ile Met Cys Ala Leu Ser 210 Arg Pro Gly Leu Leu His Gln Gly Arg Gln Arg Arg Val Arg Ala Met

230

245

Gln Leu Leu Thr Val Leu Ile Ile Phe Leu Val Cys Phe Thr Pro

250

35

	Phe His Ala Arg Gln Val Ala Val Ala Leu Trp Pro Asp Met Pro His 260 265 270	
	His Thr Ser Leu Val Val Tyr His Val Ala Val Thr Leu Ser Ser Leu 275 280 285	
5	Asn Ser Cys Met Asp Pro Ile Val Tyr Cys Phe Val Thr Ser Gly Phe 290 295 300	
	Gln Ala Thr Val Arg Gly Leu Phe Gly Gln His Gly Glu Arg Glu Pro 305 310 315 320	
10	Ser Ser Gly Asp Val Val Ser Met His Arg Ser Ser Lys Gly Ser Gly 325 330 335	
	Arg His His Ile Leu Ser Ala Gly Pro His Ala Leu Thr Gln Ala Leu 340 345 350	
	Ala Asn Gly Pro Glu Ala 355	
15	(42) INFORMATION FOR SEQ ID NO:41:	•
20	(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 30 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single	
20	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	•
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:	
	GAGAATTCAC TCCTGAGCTC AAGATGAACT	3 (
	(43) INFORMATION FOR SEQ ID NO:42:	
25	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
30	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:	
	CGGGATCCCC GTAACTGAGC CACTTCAGAT	30
	(44) INFORMATION FOR SEQ ID NO:43:	
35	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1050 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: gingle	

- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

	ATGAACTCCA	CCTTGGATGG	TAATCAGAGC	AGCCACCCTT	TTTGCCTCTT	GGCATTTGGC	60
5	TATTTGGAAA	CTGTCAATTT	TTGCCTTTTG	GAAGTATTGA	TTATTGTCTT	TCTAACTGTA	120
	TTGATTATTT	CTGGCAACAT	CATTGTGATT	TTTGTATTTC	ACTGTGCACC	TTTGTTGAAC	180
	CATCACACTA	CAAGTTATTT	TATCCAGACT	ATGGCATATG	CTGACCTTTT	TGTTGGGGTG	240
	AGCTGCGTGG	TCCCTTCTTT	ATCACTCCTC	CATCACCCCC	TTCCAGTAGA	GGAGTCCTTG	300
	ACTTGCCAGA	TATTTGGTTT	TGTAGTATCA	GTTCTGAAGA	GCGTCTCCAT	GGCTTCTCTG	360
10	GCCTGTATCA	GCATTGATAG	ATACATTGCC	ATTACTAAAC	CTTTAACCTA	TAATACTCTG	420
	GTTACACCCT	GGAGACTACG	CCTGTGTATT	TTCCTGATTT	GGCTATACTC	GACCCTGGTC	480
	TTCCTGCCTT	CCTTTTTCCA	CTGGGGCAAA	CCTGGATATC	ATGGAGATGT	GTTTCAGTGG	540
	TGTGCGGAGT	CCTGGCACAC	CGACTCCTAC	TTCACCCTGT	TCATCGTGAT	GATGTTATAT	600
	GCCCCAGCAG	CCCTTATTGT	CTGCTTCACC	TATTTCAACA	TCTTCCGCAT	CTGCCAACAG	660
15	CACACAAAGG	ATATCAGCGA	AAGGCAAGCC	CGCTTCAGCA	GCCAGAGTGG	GGAGACTGGG	720
	GAAGTGCAGG	CCTGTCCTGA	TAAGCGCTAT	GCCATGGTCC	TGTTTCGAAT	CACTAGTGTA	780
	TTTTACATCC	TCTGGTTGCC	ATATATCATC	TACTTCTTGT	TGGAAAGCTC	CACTGGCCAC	840
	AGCAACCGCT	TCGCATCCTT	CTTGACCACC	TGGCTTGCTA	TTAGTAACAG	TTTCTGCAAC	900
	TGTGTAATTI	ATAGTCTCTC	CAACAGTGTA	TTCCAAAGAG	GACTAAAGCG	CCTCTCAGGG	960
20	GCTATGTGTA	CTTCTTGTGC	AAGTCAGACT	ACAGCCAACG	ACCCTTACAC	AGTTAGAAGC	1020
	<u> እ</u> አልርርርርር ፓር	TTAATGGATG	TCATATCTGA	1			1050

(45) INFORMATION FOR SEQ ID NO:44:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 349 amino acids
- (B) TYPE: amino acid

- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:
- 30 Met Asn Ser Thr Leu Asp Gly Asn Gln Ser Ser His Pro Phe Cys Leu

		1				5				*	10		* * *	. •		15	
		Leu	Ala	Phe	Gly 20	Tyr	Leu	Glu	Thr	Val 25	. Asr	Phe	Cys	s Leu	Leu 30	ı Glu	val
5	•	Leu	Ile	Ile 35	Val	Phe	Leu	Thr	Val	Leu	ıle	·Ile	Ser	Gly 45	Asn 	Ile	lle
		Val	Ile 50	Phe	Val	Phe	His	Cys 55	Ala	Pro	Leu	Leu	Asn 60	His	His	Thr	Thr
		Ser 65	туг	Phe	Ile	Gln	Thr 70	Met	Ala	Tyr	Ala	Asp 75	Leu	Phe	Val	Gly	Val 80
10		Ser	Cys	Val	Val	Pro 85	Ser	Leu	Ser	Leu	Leu 90	His	His	Pro	Leu	Pro 95	Val
		Glu	Glu	Ser	Leu 100	Thr	Cys	Gln	Ile	Phe 105	Gly	Phe	Val	Val	Ser 110		Leu
15		Lys	Ser	Val 115	Ser	Met	Ala	Ser	Leu 120	Ala	Cys	Ile	Ser	Ile 125		Arg	Tyr
		Ile	Ala 130	Ile	Thr	Lys	Pro	Leu 135	Thr	Tyr	Asn	Thr	Leu 140	Val	Thr	Pro	Trp
	•	Arg 145	Leu	Arg	Leu	Çys	Ile 150	Phe	Leu	Ile	Trp	Leu 155	Tyr	Ser	Thr	Leu	Val 160
20		Phe	Leu	Pro	Ser	Phe 165	Phe	His	Trp	Gly	Lys 170	Pro	Gly	Tyr	His	Gly 175	Asp
		Val	Phe	Gln	Trp 180	Cys	Ala	Glu	Ser	Trp 185	His	Thr	Asp	Ser	Tyr 190	Phe	Thr
25		Leu	Phe	Ile 195	Val	Met	Met	Leu	Tyr 200	Ala	Pro	Ala	Ala	Leu 205	Ile	Val	Cys
		Phe	Thr 210	Tyr	Phe	Asn	Ile	Phe 215	Arg	Ile	Cys	Gln	Gln 220	His	Thr	Lys	Asp
		Ile 225	Ser	Glu	Arg	Gln	Ala 230	Arg	Phe	Ser	Ser	Gln 235	Ser	Gly	Glu	Thr	Gly 240
30		Glu	Val	Gln	Ala	Cys 245	Pro	Asp	Lys	Arg	Tyr 250	Ala	Met	Val	Leu	Phe 255	Arg
		Ile	Thr	Ser	Val 260	Phe	Tyr	Ile	Leu	Trp 265	Leu	Pro	Tyr	Ile	Ile 270	Tyr	Phe
35		Leu	Leu	Glu 275	Ser	Ser	Thr	Gly	His 280	Ser	Asn	Arg	Phe	Ala 285	Ser	Phe	Leu
		Thr	Thr 290	Trp	Leu	Ala		Ser 295	Asn	Ser	Phe	Cys	Asn 300	Cys	Val	Ile	Tyr

	Ser Leu Ser Asn Ser Val Phe Gln Arg Gly Leu Lys Arg Leu Ser Gly 305 310 315 320	
	Ala Met Cys Thr Ser Cys Ala Ser Gln Thr Thr Ala Asn Asp Pro Tyr 325 330 335	
5	Thr Val Arg Ser Lys Gly Pro Leu Asn Gly Cys His Ile 340 345	
	(46) INFORMATION FOR SEQ ID NO:45:	
10	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:	
15	TCCCCCGGGA AAAAAACCAA CTGCTCCAAA	30
	(47) INFORMATION FOR SEQ ID NO:46:	
20	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:	
	TAGGATCCAT TTGAATGTGG ATTTGGTGAA A	31
25	(48) INFORMATION FOR SEQ ID NO:47:	
30	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1302 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:	
	ATGTGTTTTT CTCCCATTCT GGAAATCAAC ATGCAGTCTG AATCTAACAT TACAGTGCGA	60
	GATGACATTG ATGACATCAA CACCAATATG TACCAACCAC TATCATATCC GTTAAGCTTT 1	20
35	CAAGTGTCTC TCACCGGATT TCTTATGTTA GAAATTGTGT TGGGACTTGG CAGCAACCTC 1	80

						TAACATTATT	
	ACAATGAATC	TTCATGTACT	TGATGTAATA	ATTTGTGTGG	GATGTATTCC	TCTAACTATA	300
	GTTATCCTTC	TGCTTTCACT	GGAGAGTAAC	ACTGCTCTCA	TTTGCTGTTT	CCATGAGGCT	360
	TGTGTATCTT	TTGCAAGTGT	CTCAACAGCA	ATCAACGTTT	TTGCTATCAC	TTTGGACAGA	420
5	TATGACATCT	CTGTAAAACC	TGCAAACCGA	ATTCTGACAA	TGGGCAGAGC	TGTAATGTTA	480
	ATGATATCCA	TTTGGATTTT	TTCTTTTTC	TCTTTCCTGA	TTCCTTTTAT	TGAGGTAAAT	540
	TTTTTCAGTC	TTCAAAGTGG	AAATACCTGG	GAAAACAAGA	CACTTTTATG	TGTCAGTACA	600
	AATGAATACT	ACACTGAACT	GGGAATGTAT	TATCACCTGT	TAGTACAGAT	CCCAATATTC	660
	TTTTTCACTG	TTGTAGTAAT	GTTAATCACA	TACACCAAAA	TACTTCAGGC	TCTTAATATT	720
10	CGAATAGGCA	CAAGATTTTC	AACAGGGCAG	AAGAAGAAAG	CAAGAAAGAA	AAAGACAATT	780
	TCTCTAACCA	CACAACATGA	GGCTACAGAC	ATGTCACAAA	GCAGTGGTGG	GAGAAATGTA	840
	GTCTTTGGTG	TAAGAACTTC	AGTTTCTGTA	ATAATTGCCC	TCCGGCGAGC	TGTGAAACGA	900
	CACCGTGAAC	GACGAGAAAG	ACAAAAGAGA	GTCTTCAGGA	TGTCTTTATT	GATTATTTCT	960
	ACATTTCTTC	TCTGCTGGAC	ACCAATTTCT	GTTTTAAATA	CCACCATTTT	ATGTTTAGGC	1020
15	CCAAGTGACC	TTTTAGTAAA	ATTAAGATTG	TGTTTTTTAG	TCATGGCTTA	TGGAACAACT	1080
	ATATTTCACC	СТСТАТТАТА	TGCATTCACT	AGACAAAAAT	TTCAAAAGGT	CTTGAAAAGT	1140
	AAAATGAAAA	AGCGAGTTGT	TTCTATAGTA	GAAGCTGATC	CCCTGCCTAA	TAATGCTGTA	1200
	ATACACAACT	CTTGGATAGA	TCCCAAAAGA	аасаааааа	TTACCTTTGA	AGATAGTGAA	1260
	ATAAGAGAAA	AACGTTTAGT	GCCTCAGGTT	GTCACAGACT	AG		1302

- 20 (49) INFORMATION FOR SEQ ID NO:48:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 433 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
- - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

Met Cys Phe Ser Pro Ile Leu Glu Ile Asn Met Gln Ser Glu Ser Asn 1 5 10 15

30 Ile Thr Val Arg Asp Asp Ile Asp Asp Ile Asn Thr Asn Met Tyr Gln 20 25 30

	Pro	Leu	ser 35	ıyr	Pro	Leu	ser	40	GIN	vai	ser	Leu	1'nr 45	GIÀ	Pne	Leu
	Met	Leu 50	Glu	Ile	Val	Leu	Gly 55	Leu	Gly	Ser	Asn	Leu 60	Thr	Val	Leu	Val
5	Leu 65	Tyr	Cys	Met	Lys	Ser 70	Asn	Leu	Ile	Asn	Ser 75	Val	Ser	Asn	Ile	Ile 80
•	Thr	Met	Asn	Leu	His 85	Val	Leu	Asp	Val	Ile 90	Ile	Cys	Val	Gly	Cys 95	Ile
10	Pro	Leu	Thr	Ile 100	Val	Ile	Leu	Leu	Leu 105	Ser	Leu	Glu	Ser	Asn 110	Thr	Ala
	Leu	Ile	Cys 115	Cys	Phe	His	Glu	Ala 120	Cys	Val	Ser	Phe	Ala 125	Ser	Val	Ser
	Thr	Ala 130	Ile	Asn	Val	Phe	Ala 135	Ile	Thr	Leu	Asp	Arg 140	Tyr	Asp	Ile	Ser
15	Val 145	Lys	Pro	Ala	Asn	Arg 150	Ile	Leu	Thr	Met	Gly 155	Arg	Ala	Val	Met	Leu 160
	Met	Ile	Ser	Ile	Trp 165	Ile	Phe	Ser	Phe	Phe 170	Ser	Phe	Leu	Ile	Pro 175	Phe
20	Ile	Glu	Val	Asn 180	Phe	Phe	Ser.	Leu	Gln 185	Ser	Gly	Asn	Thr	Trp 190	Glu	Asn
	Lys	Thr	Leu 195	Leu	Cys	Val	Ser	Thr 200	Asn	Glu	Tyr	Tyr	Thr 205	Glu	Leu	Gly
	Met	Tyr 210	Tyr	His	Leu	Leu	Val 215	Gln	Ile	Pro	Ile	Phe 220	Phe	Phe	Thr	Val
25	Val 225	Val	Met	Leu	Ile	Thr 230	Tyr	Thr	Lys	Ile	Leu 235	Gln	Ala	Leu	Asn	Ile 240
	Arg	Ile	Gly	Thr	Arg 245	Phe	Ser	Thr	Gly	Gln 250	Lys	Lys	Lys	Ala	Arg 255	Lys
30	Lys	Lys	Thr	Ile 260	Ser	Leu	Thr	Thr	Gln 265	His	Glu	Ala	Thr	Asp 270	Met	Ser
	Gln	Ser	Ser 275	Gly	Gly	Arg	Asn	Val 280	Val	Phe	Gly	Val	Arg 285	Thr	Ser	Val
	Ser	Val 290	Ile	Ile	Ala	Leu	Arg 295	Arg	Ala	Val	Lys	Arg 300	His	Arg	Glu	Arg
35	Arg 305	Glu	Arg	Gln	Lys	Arg 310	Val	Phe	Arg	Met	Ser 315	Leu	Leu	Ile	Ile	Ser 320
	Thr	Phe	Leu	Leu	Cys	Trp	Thr	Pro	Ile	Ser	Val	Leu	Asn	Thr	Thr	Ile

WO 00/22129 PCT/US99/23938

39

	•				•	325				. <i>.</i>	330			•	s.	335		
		Leu	Cys	Leu	Gly 340	Pro	Ser	Asp	Leu	Leu 345	Val	Lys	Leu	Arg	Leu 350	Cys	Phe	
5		Leu	Val	Met 355	Ala	Tyr	Gly	Thr	Thr 360	Ile	Phe	His	Pro	Leu 365	Leu	Tyr	Ala	
		Phe	Thr 370	Arg	Gln	Lys	Phe	Gln 375	Lys	Val	Leu	Lys	Ser 380	Lys	Met	Lys	Lys	
		Arg 385	Val	Val	Ser	Ile	Val 390	Glu	Ala	Asp	Pro	Leu 395	Pro	Asn	Asn	Ala	Val 400	
10		Ile	His	Asn	Ser	Trp 405	Ile	Asp	Pro	Lys	Arg 410	Asn	Lys	Lys	Ile	Thr 415	Phe	
		Glu	Asp	Ser	Glu 420	Ile	Arg	Glu	Lys	Arg 425	Leu	Val	Pro	Gln	Val 430	Val	Thr	
15		Asp																٠.
	(50)	INFO	RMAT	rion	FOR	SEQ	ID 1	10:49):									
	•	(i)	SEÇ	UENC	E CH	IARAC	TER	STIC	S:									
20	 (50) INFORMATION FOR SEQ ID NO:49: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 																	
	,	(ii	.) MC	LECU	LE I	YPE:	DNA	ı (ge	nomi	.c)								
		(xi) SE	QUEN	ICE I	ESCR	IPTI	ON:	SEO	ID N	IO : 4 9) <u>:</u>						
	GTGAZ			٠					_									30
25	(51)			ION					١.	•								3.0
30	,		SEQ (A) (B) (C)	UENC LEN TYP STR	E CH GTH: E: n	ARAC 31 ucle	TERI base ic a	STIC pai cid ingl	S: rs				-					
		(ii	.) MC	LECU	LE T	YPE:	DNA	(ge	nomi	c)								
		(xi) SE	QUEN	CE D	ESCR	IPTI	ON:	SEQ	ID N	0:50	:						
	GCAGA	ATTC	C CG	GTGG	CGTG	TTG	TGGT	GCC	С									31
	(52)	INFO	RMAT	'ION	FOR	SEQ	ID N	10:51	:									
35		(i)	SEQ	UENC	Е СН	ARAC	TERI	STIC	s:									

(A) LENGTH: 1209 base pairs

- (B) TYPE: nucleic acid(C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

ATGTTGTGTC CTTCCAAGAC AGATGGCTCA GGGCACTCTG GTAGGATTCA CCAGGAAACT CATGGAGAAG GGAAAAGGGA CAAGATTAGC AACAGTGAAG GGAGGGAGAA TGGTGGGAGA GGATTCCAGA TGAACGGTGG GTCGCTGGAG GCTGAGCATG CCAGCAGGAT GTCAGTTCTC AGAGCAAAGC CCATGTCAAA CAGCCAACGC TTGCTCCTTC TGTCCCCAGG ATCACCTCCT 240 10 CGCACGGGGA GCATCTCCTA CATCAACATC ATCATGCCTT CGGTGTTCGG CACCATCTGC 300 CTCCTGGGCA TCATCGGGAA CTCCACGGTC ATCTTCGCGG TCGTGAAGAA GTCCAAGCTG 360 CACTGGTGCA ACAACGTCCC CGACATCTTC ATCATCAACC TCTCGGTAGT AGATCTCCTC TTTCTCCTGG GCATGCCCTT CATGATCCAC CAGCTCATGG GCAATGGGGT GTGGCACTTT 480 GGGGAGACCA TGTGCACCCT CATCACGGCC ATGGATGCCA ATAGTCAGTT CACCAGCACC 540 15 TACATCCTGA CCGCCATGGC CATTGACCGC TACCTGGCCA CTGTCCACCC CATCTCTTCC 600 ACGAAGTTCC GGAAGCCCTC TGTGGCCACC CTGGTGATCT GCCTCCTGTG GGCCCTCTCC TTCATCAGCA TCACCCCTGT GTGGCTGTAT GCCAGACTCA TCCCCTTCCC AGGAGGTGCA 720 GTGGGCTGCG GCATACGCCT GCCCAACCCA GACACTGACC TCTACTGGTT CACCCTGTAC 780 CAGTTTTTCC TGGCCTTTGC CCTGCCTTTT GTGGTCATCA CAGCCGCATA CGTGAGGATC 20 CTGCAGCGCA TGACGTCCTC AGTGGCCCCC GCCTCCCAGC GCAGCATCCG GCTGCGGACA AAGAGGGTGA CCCGCACAGC CATCGCCATC TGTCTGGTCT TCTTTGTGTG CTGGGCACCC 960 TACTATGTGC TACAGCTGAC CCAGTTGTCC ATCAGCCGCC CGACCCTCAC CTTTGTCTAC 1020 TTATACAATG CGGCCATCAG CTTGGGCTAT GCCAACAGCT GCCTCAACCC CTTTGTGTAC 1080 ATCGTGCTCT GTGAGACGTT CCGCAAACGC TTGGTCCTGT CGGTGAAGCC TGCAGCCCAG GGGCAGCTTC GCGCTGTCAG CAACGCTCAG ACGCTGACG AGGAGAGGAC AGAAAGCAAA 1200 GGCACCTGA 1209

(53) INFORMATION FOR SEQ ID NO:52:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 402 amino acids
- 30 (B) TYPE: amino acid

-

- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

	(X)	L) 51	-QOE	ACE I	JESCI	KIPI.	LOIN:	SEQ	ו עו	NO:5	۷:					
5	Met 1	Leu	Cys	Pro	Ser 5	Lys	Thr	Asp	Gly	Ser 10	Gly	His	Ser	Gly	Arg 15	Ile
	His	Gln	Glu	Thr 20	His	Gly	Glu	Gly	Lys 25	Arg	Asp	Lys	Ile	Ser 30	Asn	Ser
10	Glu	Gly	Arg 35	Glu	Asn	Gly	Gly	Arg 40	Gly	Phe	Gln	Met	Asn 45	Gly	Gly	Ser
	 Leu	Glu 50	Ala	Glu	His	Ala	Ser 55	Arg	Met	Ser	Val	Leu 60	Arg	Ala	Lys	Pro
	Met 65	Ser	Asn	Ser	Gln	Arg 70	Leu	Leu	Leu	Leu	Ser 75	Pro	Gly	Ser	Pro	Pro 80
15	Arg	Thr	Gly	Ser	Ile 85	Ser	Tyr	Ile	Asn	Ile 90	Ile	Met	Pro	Ser	Val 95	Phe
	 Gly	Thr	Ile	Cys 100	Leu	Leu	Gly	Ile	Ile 105	Gly	Asn	Ser	Thr	Val 110	Ile	Phe
20	Ala	Val	Val 115	Lys	Lys	Ser	Lys	Leu 120	His	Trp	Cys	Asn	Asn 125	Val	Pro	Asp :
	Ile	Phe 130	Ile	Ile	Asn	Leu	Ser 135	Val	Val	Asp	Leu	Leu 140	Phe	Leu	Leu	Gly
	Met 145	Pro	Phe	Met	Ile	His 150		Leu	Met	Gly	Asn 155	Gly	Val	Trp	His	Phe 160
25	Gly	Glu	Thr	Met	Cys 165	Thr	Leu	Ile	Thr	Ala 170	Met	Asp	Ala	Asn	Ser 175	Gln
	Phe	Thr	Ser	Thr 180	Tyr	Ile	Leu	Thr	Ala 185	Met	Ala	Ile	Asp	Arg 190	Tyr	Leu
30	Ala	Thr	Val 195	His	Pro	Ile	Ser	Ser 200	Thr	Lys	Phe	Arg	Lys 205	Pro	Ser	Val
	Ala	Thr 210	Leu	Val	Ile	Cys	Leu 215	Leu	Trp	Ala	Leu	Ser 220	Phe	Ile	Ser	Ile
	Thr 225	Pro	Val	Trp	Leu	туr 230	Ala	Arg	Leu	Ile	Pro 235	Phe	Pro	Gly	Gly	Ala 240
35	Val	Gly	Cys	Gly	Ile 245	Arg	Leu	Pro	Asn	Pro 250	Asp	Thr	Asp	Leu	Tyr 255	Trp

							,	4	2							
	Ph	e Thr	Leu	Tyr 260	Gln	Phe	Phe	Leu	Ala 265	Phe	Ala	Leu	Pro	Phe 270	Val	Val
	Il	e Thr	Ala 275	Ala	Tyr	Val	Arg	Ile 280	Leu	Gln	Arg	Met	Thr 285	Ser	Ser	Val
. 5	Al	a Pro 290		Ser	Gln	Arg	Ser 295	Ile	Arg	Leu	Arg	Thr 300	Lys	Arg	Val	Thr
	Ar 30	g Thr 5	Ala	Ile	Ala	Ile 310		Leu	Val	Phe	Phe	Val	Cys	Trp	Ala	Pro 320
10	ту	r Tyr	Val	Leu	Gln 325	Leu	Thr	Gln	Leu	Ser 330	Ile	Ser	Arg	Pro	Thr 335	Leu
	Th	r Phe	Val	Tyr 340	Leu	Tyr	Asn	Ala	Ala 345	Ile	Ser	Leu	Gly	Tyr 350		Asn
	Se	r Cys	Leu 355	Asn	Pro	Phe	Val	Tyr 360	Ile	Val	Leu	Cys	Glu 365	Thr	Phe	Arg
15	Ly	s Arg 370		Val	Leu	Ser	Val 375	Lys	Pro	Ala	Ala	Gln 380	Gly	Gln	Leu	Arg
	Al 38	a Val	Ser	Asn	Ala	Gln 390		Ala	Asp	Glu	Glu 395	Arg	Thr	Glu	Ser	Lys 400
20	Gl (54) IN	y Thr		FOR	SEQ	ID :	NO:5	3:								
25		(B) LEI) TY:) STI) TO:	NGTH PE: 1 RAND: POLO	: 27 nucle EDNE GY:	bas eic SS: line	e pa acid sing ar	irs le	ic)						,	
,) GGCGGAT	xi) S						SEQ	ID 1	NO:5	3:					27
••	(55) IN				•			4:								21
35	((B	QUEN LE TY TY TO	NGTH PE: RAND	: 27 nucl EDNE	bas eic SS:	e pa acid sing	irs								•
2.5	(ii) M						enom	ic)			٠				

PCT/US99/23938

5

30

GGCGGATCCC TACACGGCAC TGCTGAA

27

(56) INFORMATION FOR SEQ ID NO:55:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1128 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:
- ATGGATGTGA CTTCCCAAGC CCGGGGCGTG GGCCTGGAGA TGTACCCAGG CACCGCGCAC 60 10 GCTGCGGCCC CCAACACCAC CTCCCCCGAG CTCAACCTGT CCCACCCGCT CCTGGGCACC 120 GCCCTGGCCA ATGGGACAGG TGAGCTCTCG GAGCACCAGC AGTACGTGAT CGGCCTGTTC 180 CTCTCGTGCC TCTACACCAT CTTCCTCTTC CCCATCGGCT TTGTGGGCAA CATCCTGATC 240 CTGGTGGTGA ACATCAGCTT CCGCGAGAAG ATGACCATCC CCGACCTGTA CTTCATCAAC 300 CTGGCGGTGG CGGACCTCAT CCTGGTGGCC GACTCCCTCA TTGAGGTGTT CAACCTGCAC 360 GAGCGGTACT ACGACATCGC CGTCCTGTGC ACCTTCATGT CGCTCTTCCT GCAGGTCAAC 420 ATGTACAGCA GCGTCTTCTT CCTCACCTGG ATGAGCTTCG ACCGCTACAT CGCCCTGGCC 480 AGGGCCATGC GCTGCAGCCT GTTCCGCACC AAGCACCACG CCCGGCTGAG CTGTGGCCTC 540 ATCTGGATGG CATCCGTGTC AGCCACGCTG GTGCCCTTCA CCGCCGTGCA CCTGCAGCAC 600 ACCGACGAGG CCTGCTTCTG TTTCGCGGAT GTCCGGGAGG TGCAGTGGCT CGAGGTCACG 660 20 CTGGGCTTCA TCGTGCCCTT CGCCATCATC GGCCTGTGCT ACTCCCTCAT TGTCCGGGTG 720 CTGGTCAGGG CGCACCGGCA CCGTGGGCTG CGGCCCCGGC GGCAGAAGGC GCTCCGCATG 780 840 GTGCACCTCC TGCAGCGGAC GCAGCCTGGG GCCGCTCCCT GCAAGCAGTC TTTCCGCCAT 900 GCCCACCCC TCACGGGCCA CATTGTCAAC CTCGCCGCCT TCTCCAACAG CTGCCTAAAC 960 25 CCCCTCATCT ACAGCTTTCT CGGGGAGACC TTCAGGGACA AGCTGAGGCT GTACATTGAG 1020 CAGAAAACAA ATTTGCCGGC CCTGAACCGC TTCTGTCACG CTGCCCTGAA GGCCGTCATT 1080 CCAGACAGCA CCGAGCAGTC GGATGTGAGG TTCAGCAGTG CCGTGTGA 1128
 - (57) INFORMATION FOR SEQ ID NO:56:
 - (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 375 amino acids

			(C)	STF	RANDI	amino EDNES GY: r	SS:		ant		-						
5		(ii) MC	LECT	JLE 1	TYPE:	pro	teir	1								
		(xi	L) SI	EQUEN	ICE I	DESCR	RIPTI	ON:	SEQ	ID 1	10 : 5€	5:	•				
		Met 1	Asp	Val	Thr	Ser 5	Gln	Ala	Arg	Gly	Val 10	Gly	Leu	Glu	Met	Tyr 15	Pro
10		Glý	Thr	Ala	His 20	Ala	Ala	Ala	Pro	Asn 25	Thr	Thr	Ser	Pro	Glu 30	Leu	Asn
		Leu	Ser	His 35	Pro	Leu	Leu	Gly	Thr 40	Ala	Leu	Ala	Asn	Gly 45	Thr	Gly	Glu
	,	Leu	Ser 50	Glu	His	Gln	Gln	Tyr 55	Val	Ile	Gly	Leu	Phe 60	Leu	Ser	Cys	Leu
15		Tyr 65	Thr	Ile	Phe	Leu	Phe 70	Pro	Ile	Gly	Phe	Val 75	Gly	Asn	Ile	Leu	Ile 80
		Leu	Val	Val	Asn	Ile 85	Ser	Phe	Arg	Glu	Lys 90	Met	Thr	Ile	Pro	Asp 95	Leu
20		Tyr	Phe	Ile	Asn 100	Leu	Ala	Val	Ala	Asp 105	Leu	Ile	Leu	Val	Ala 110	Asp	Ser
		Leu	Ile	Glu 115	Val	Phe	Asn	Leu	His 120	Glu	Arg	Tyr	Tyr	Asp 125	Ile	Ala	Val
		Leu	Cys 130	Thr	Phe	Met	Ser	Leu 135	Phe	Leu	Gln	Val	Asn 140	Met	Tyr	Ser	Ser
25		Val 145	Phe	Phe	Leu	Thr	Trp 150	Met	Ser	Phe	Asp	Arg 155	Tyr	Ile	Ala	Leu	Ala 160
		Arg	Ala	Met	Arg	Cys 165	Ser	Leu	Phe	Arg	Thr 170	Lys	His	His	Ala	Arg 175	Lev
30		Ser	Cys	Gly	Leu 180	Ile	Trp	Met	Ala	Ser 185	Val	Ser	Ala	Thr	Leu 190	Val	Pro
		Phe	Thr	Ala 195	Val	His	Leu	Gln	His 200	Thr	Asp	Glu	Ala	Cys 205	Phe	Cys	Phe
		Ala	Asp 210	Val	Arg	Glu	Val	Gln 215	Trp	Leu	Glu	Val	Thr 220	Leu	Gly	Phe	Ile
35		Val 225	Pro	Phe	Ala	Ile	Ile 230	Gly	Leu	Cys	Tyr	Ser 235	Leu	Ile	Val	Arg	Va]
		Leu	Val	Arg	Ala	His	Arg	His	Arg	Gly	Leu	Arg	Pro	Arg	Arg	Gln	Lys

PCT/US99/23938

						245				•	250					255		
•		Ala	Leu	Arg	Met 260	Ile	Leu	Ala	Val	Val 265	Leu	Val	Phe	Phe	Val 270	Cys	Trp	
5		Leu	Pro	Glu 275	Asn	Val	Phe	Ile	Ser 280	Val	His	Leu	Leu	Gln 285	Arg	Thr	Gln	
		Pro	Gly 290	Ala	Ala	Pro	Cys	Lys 295	Gln	Ser	Phe	Arg	His 300	Ala	His	Pro	Leu	
		Thr 305	Gly	His	Ile	Val	Asn 310	Leu	Ala	Ala	Phe	Ser 315	Asn	Ser	Cys	Leu	Asn 320	
10		Pro	Leu	Ile	Tyr	Ser 325	Phe	Leu	Gly	Glu	Thr 330	Phe	Arg	Asp	Lys	Leu 335	Arg	
	, · · · ·	Leu	Tyr	Ile	Glu 340	Gln	Lys	Thr	Asn	Leu 345	Pro	Ala	Leu		Arg 350	Phe	Cys	,
15		His	Ala	Ala 355	Leu	Lys	Ala	Val	Ile 360	Pro	Asp	Ser	Thr	Glu 365	Gln	Ser	Asp ."	
		Val	Arg 370	Phe	Ser	Ser	Ala	Val 375			,		-				·	
	(58)	INF	ORMA'	rion	FOR	SEQ	ID 1	NO:5	7:									
20			(A) (B) (C)	LEI TYI	NGTH PE: 1	: 31 nucle EDNES	base eic a	ISTIC e pa: acid sing! ar	irs	. ,	:	. •						
		(i:	i) Mo	OLEC	JLE T	TYPE:	: DNZ	A (ge	enom:	ic)								
25		(x:	i) SI	EOUEI	NCE I	DESCI	RIPT:	ION:	SEQ	ID 1	NO:51	7:	•		-			
	AAGG	AATTO	CA CO	GCC	GGT	G ATO	GCCA!	TTCC	C								3	31
	(59)	INFO	ORMA'	TION	FOR	SEO	ID 1	NO : 58	3:									
30			SE(QUEN(CE CI	HARAG	CTER:	ISTIC e pa:	CS:	,		<i>:</i> •			. •			
					RANDI POLOC			sing] ar	le									
		(i:	i) Mo	OLECT	ULE 1	TYPE:	: DNZ	A (ge	enom	ic)					•			
		(x:	i) sı	EQUEI	NCE I	DESCI	RIPT	ION:	SEQ	ID 1	NO:58	В:						
35	GGTG	GATC	CA T	AAAC	ACGG	G CG	TTGAG	GGAC							•		3	30
	(60)	INF	ORMA!	rion	FOR	SEQ	ID I	NO : 5	9:									

46 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 960 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (genomic) (xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

5

- ATGCCATTCC CAAACTGCTC AGCCCCCAGC ACTGTGGTGG CCACAGCTGT GGGTGTCTTG CTGGGGCTGG AGTGTGGGCT GGGTCTGCTG GGCAACGCGG TGGCGCTGTG GACCTTCCTG 10 TTCCGGGTCA GGGTGTGGAA GCCGTACGCT GTCTACCTGC TCAACCTGGC CCTGGCTGAC CTGCTGTTGG CTGCGTGCCT GCCTTTCCTG GCCGCCTTCT ACCTGAGCCT CCAGGCTTGG CATCTGGGCC GTGTGGGCTG CTGGGCCCTG CGCTTCCTGC TGGACCTCAG CCGCAGCGTG 300 GGGATGGCCT TCCTGGCCGC CGTGGCTTTG GACCGGTACC TCCGTGTGGT CCACCCTCGG CTTAAGGTCA ACCTGCTGTC TCCTCAGGCG GCCCTGGGGG TCTCGGGCCT CGTCTGGCTC 15 CTGATGGTCG CCCTCACCTG CCCGGGCTTG CTCATCTCTG AGGCCGCCCA GAACTCCACC 480 AGGTGCCACA GTTTCTACTC CAGGGCAGAC GGCTCCTTCA GCATCATCTG GCAGGAAGCA 540 CTCTCCTGCC TTCAGTTTGT CCTCCCCTTT GGCCTCATCG TGTTCTGCAA TGCAGGCATC 600 ATCAGGGCTC TCCAGAAAAG ACTCCGGGAG CCTGAGAAAC AGCCCAAGCT TCAGCGGGCC CAGGCACTGG TCACCTTGGT GGTGGTGCTG TTTGCTCTGT GCTTTCTGCC CTGCTTCCTG 720 20 GCCAGAGTCC TGATGCACAT CTTCCAGAAT CTGGGGAGCT GCAGGGCCCT TTGTGCAGTG 780 GCTCATACCT CGGATGTCAC GGGCAGCCTC ACCTACCTGC ACAGTGTCGT CAACCCCGTG 840 GTATACTGCT TCTCCAGCCC CACCTTCAGG AGCTCCTATC GGAGGGTCTT CCACACCCTC CGAGGCAAAG GGCAGGCAGC AGAGCCCCCA GATTTCAACC CCAGAGACTC CTATTCCTGA (61) INFORMATION FOR SEQ ID NO:60:
- 25 (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 319 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
- 30 (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

Met Pro Phe Pro Asn Cys Ser Ala Pro Ser Thr Val Val Ala Thr Ala

	1				5					10	•	٠			15	
	Val	Gly	Val	Leu 20	Leu	Gly	Leu	Glu	Cys 25	Gly	Leu	Gly	Leu	Leu 30	Gly	Asn
5	Ala	Val	Ala 35	Leu	Trp	Thr	Phe	Leu 40	Phe	Arg	Val	Arg	Val 45	Trp	Lys	Pro
	Tyr	Ala 50	Val	Tyr	Leu	Leu	Asn 55	Leu	Ala	Leu	Ala	Asp 60	Leu	Leu	Leu	Ala
	Ala 65	Cys	Leu	Pro	Phe	Leu 70	Ala	Ala	Phe	Tyr	Leu 75	Ser	Leu	Gln	Ala	Trp 80
10	His	Leu	Gly	Arg	Val 85	Gly	Cys	Trp	Ala	Leu 90	Arg	Phe	Leu	Leu	Asp 95	Leu
	Ser	Arg	Ser	Val 100	Gly	Met	Ala	Phe	Leu 105	Ala	Ala	Val	Ala	Leu 110	Asp	Arg
15	Tyr	Leu	Arg 115	Val	Val	His	Pro	Arg 120	Leu	Lys	Val	Asn	Leu 125	Leu	Ser	Pro
	Gln	Ala 130	Ala	Leu	Gly	Val	Ser 135	Gly	Leu	Val	Trp	Leu 140	Leu	Met	Val	Ala
	Leu 145	Thr	Cys	Pro	Gly	Leu 150	Leu	Ile	Ser	Glu	Ala 155	Ala	Gln	Asn	Ser	Thr 160
20	Arg	Cys	His	Ser	Phe 165	Tyr	Ser	Arg	Ala	Asp 170	Gly	Ser	Phe	Ser	Ile 175	Ile
	Trp	Gln	Glu	Ala 180	Leu	Ser	Cys	Leu	Gln 185	Phe	Val	Leu	Pro	Phe 190	Gly	Leu
25	Ile	Val	Phe 195	Cys	Asn	Ala	Gly	Ile 200	Ile	Arg	Ala	Leu	Gln 205	Lys	Arg	Leu
	Arg	Glu 210	Pro	Glu	Lys	Gln	Pro 215	Lys	Leu	Gln	Arg	Ala 220	Gln	Ala	Leu	Val
	Thr 225	Leu	Val	Val	Val	Leu 230	Phe	Ala	Leu	Cys	Phe 235	Leu	Pro	Cys	Phe	Leu 240
30	Ala	Arg	Val	Leu	Met 245		Ile	Phe	Gln	Asn 250		·Gly	Ser	Cys	Arg 255	Ala
	Leu	Cys	Ala	Val 260		His	Thr	Ser	Asp 265		Thr	Gly	Ser	Leu 270		Tyr
35	Leu	His	Ser 275		Val	Asn	Pro	Val 280		Tyr	Cys	Phe	Ser 285		Pro	Thr
	Phe	Arg		Ser	Tyr	Arg	Arg 295		Phe	His		Leu 300		Gly	Lys	Gly

Gln Ala Ala Glu Pro Pro Asp Phe Asn Pro Arg Asp Ser Tyr Ser 305 310 315

(62) INFORMATION FOR SEQ ID NO:61:

5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1143 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

ATGGAGGAAG GTGGTGATTT TGACAACTAC TATGGGGCAG ACAACCAGTC TGAGTGTGAG TACACAGACT GGAAATCCTC GGGGGCCCTC ATCCCTGCCA TCTACATGTT GGTCTTCCTC CTGGGCACCA CGGGAAACGG TCTGGTGCTC TGGACCGTGT TTCGGAGCAG CCGGGAGAAG ACGCTGCCCC TGTGGGCTAC CTACACGTAC CGGGACTATG ACTGGCCCTT TGGGACCTTC 300 TTCTGCAAGC TCAGCAGCTA CCTCATCTTC GTCAACATGT ACGCCAGCGT CTTCTGCCTC 360 ACCGGCCTCA GCTTCGACCG CTACCTGGCC ATCGTGAGGC CAGTGGCCAA TGCTCGGCTG 420 AGGCTGCGGG TCAGCGGGGC CGTGGCCACG GCAGTTCTTT GGGTGCTGGC CGCCCTCCTG GCCATGCCTG TCATGGTGTT ACGCACCACC GGGGACTTGG AGAACACCAC TAAGGTGCAG 540 20 TGCTACATGG ACTACTCCAT GGTGGCCACT GTGAGCTCAG AGTGGGCCTG GGAGGTGGGC 600 CTTGGGGTCT CGTCCACCAC CGTGGGCTTT GTGGTGCCCT TCACCATCAT GCTGACCTGT TACTTCTTCA TCGCCCAAAC CATCGCTGGC CACTTCCGCA AGGAACGCAT CGAGGGCCTG CGGAAGCGGC GCCGGCTGCT CAGCATCATC GTGGTGCTGG TGGTGACCTT TGCCCTGTGC TGGATGCCCT ACCACCTGGT GAAGACGCTG TACATGCTGG GCAGCCTGCT GCACTGGCCC TGTGACTTTG ACCTCTTCCT CATGAACATC TTCCCCTACT GCACCTGCAT CAGCTACGTC 900 AACAGCTGCC TCAACCCCTT CCTCTATGCC TTTTTCGACC CCCGCTTCCG CCAGGCCTGC 960 ACCTCCATGC TCTGCTGTGG CCAGAGCAGG TGCGCAGGGA CCTCCCACAG CAGCAGTGGG 1020 GAGAAGTCAG CCAGCTACTC TTCGGGGCAC AGCCAGGGGC CCGGCCCCAA CATGGGCAAG 1080 GGTGGAGAAC AGATGCACGA GAAATCCATC CCCTACAGCC AGGAGACCCT TGTGGTTGAC 1140 30 TAG 1143

5

(63) INFORMATION FOR SEQ ID NO:62:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 380 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

	(X)	L) 51	SQUEI	NCE I	DESCI	KIPT.	LOIN:	SEQ	ו עו	NO: 6.	3:					
10	Met 1	Glu	Glu	Gly	Gly 5	Asp	Phe	Asp	Asn	Туг 10	Tyr	Gly	Ala	Asp	Asn 15	Gln
	Ser	Glu	Cys	Glu 20	Tyr	Thr	Asp	Trp	Lys 25	Ser	Ser	Gly	Ala	Leu 30	Ile	Pro
	Ala	Ile	Tyr 35	Met	Leu	Val	Phe	Leu 40	Leu	Gly	Thr	Thr	Gly 45	Asn	Gly	Leu
15	Val	Leu 50	Trp	Thr	Val	Phe	Arg 55	Ser	Ser	Arg	Glu	Lys 60	Arg	Arg	Ser	Ala
·	Asp 65	Ile	Phe	Ile	Ala	Ser 70	Leu	Ala	Val	Ala	Asp 75	Leu	Thr	Phe	Val	Val 80
20	Thr	Leu	Pro	Leu	Trp 85	Ala	Thr	Tyr	Thr	Tyr 90	Arg	Asp	Tyr	Asp	Trp 95	Pro
	Phe	Gly	Thr	Phe 100	Phe	Cys	Lys	Leu	Ser 105	Ser	Tyr	Leu	Ile	Phe 110	Val	Asn
	Met	Tyr	Ala 115	Ser	Val	Phe	Cys	Leu 120	Thr	Gly	Leu	Ser	Phe 125	Asp	Arg	Tyr
25	Leu	Ala 130	Ile	Val	Arg	Pro	Val 135	Ala	Asn	Ala	Arg	Leu 140	Arg	Leu	Arg	Val
	Ser 145	Gly	Ala	Val	Ala	Thr 150	Ala	Val	Leu	Trp	Val 155	Leu	Ala	Ala	Leu	Leu 160
30	Ala	Met	Pro	Val	Met 165	Val	Leu	Arg	Thr	Thr 170	Gly	Asp	Leu	Glu	Asn 175	Thr
	Thr	Lys	Val	Gln 180	Cys	Tyr	Met	Asp	Tyr 185	Ser	Met	Val	Ala	Thr 190	Val	Ser
	Ser	Glu	Trp 195	Ala	Trp	Glu	Val	Gly 200	Leu	Gly	Val	Ser	Ser 205	Thr	Thr	Val
35	Gly	Phe 210	Val	Val	Pro	Phe	Thr 215	Ile	Met	Leu	Thr	Cys 220	Tyr	Phe	Phe	Ile

Ala Gln Thr Ile Ala Gly His Phe Arg Lys Glu Arg Ile Glu Gly Leu

	2.	25					230					235					240	
	Az	rg	Lys	Arg	Arg	Arg 245	Leu	Leu	Ser	Ile	Ile 250		Val	Leu	Val	Val 255	Thr	
5	Pł	ne .	Ala	Leu	Cys 260	Trp	Met	Pro	Tyr	His 265	Leu	Val	Lys	Thr	Leu 270	Tyr	Met	
	Le	eu (Gly	Ser 275	Leu	Leu	His	Trp	Pro 280	Cys	Asp	Phe	Asp	Leu 285	Phe	Leu	Met	
	As	n :	lle 290	Phe	Pro	Tyr	Cys	Thr 295	Cys	Ile	Ser	Tyr	Val 300	Asn	Ser	Cys	Leu	
10	As	n :	Pro	Phe	Leu	Tyr	Ala 310	Phe	Phe	Asp	Pro	Arg 315	Phe	Arg	Gln	Ala	Cys 320	
	Th		Ser	Met	Leu	Cys 325	Cys	Gly	Gln		Arg 330	Cys	Ala	Gly	Thr	Ser 335	His	
15					340		Lys			345					350			
	Gl	уІ	Pro	Gly 355	Pro	Asn	Met	Gly	Lys 360	Gly	Gly	Glu	Gln	Met 365	His	Glu	_	
•		3	370				Gln	375		Leu	Val	Val	Asp 380				j	
20	(64) IN	FOF	TAMS	'ION	FOR	SEQ	ID N	10:63	:									
	(i)	(A) (B)	LEN TYP	GTH: E: n	31 ucle	TERI base	pai cid	rs		• .							
25							S: s inea		e					•				
-							DNA							. : .				
	(:	xi)	SE	QUEN	CE D	ESCR	IPTI	ON:	SEQ	ID N	0:63	:						
٠	TGAGAAT'	rct	' GG	TGAC	TCAC	AGC	CGGC	ACA	G								. :	31
	(65) IN	FOR	TAM	ION	FOR .	SEQ	ID N	0:64	:	٠		•	•		•			
30			(A) (B) (C)	LENG TYP: STR	GTH: E: n ANDE	31 : ucle DNES	TERI base ic a S: s inea	pai cid ingl	rs								٠	
35	£)	ii)	MO	LECU	LE T	YPE:	DNA	(ge	nomi	c)			٠					
	()	(i)	SE	QUEN	CE DI	ESCR:	IPTI	ON:	SEQ :	ID N	0:64	:						

5

30

GCCGGATCCA AGGAAAAGCA GCAATAAAAG G

31

(66) INFORMATION FOR SEQ ID NO:65:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1119 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

10	ATGAACTACC	CGCTAACGCT	GGAAATGGAC	CTCGAGAACC	TGGAGGACCT	GTTCTGGGAA	60
	CTGGACAGAT	TGGACAACTA	TAACGACACC	TCCCTGGTGG	AAAATCATCT	CTGCCCTGCC	120
	ACAGAGGGTC	CCCTCATGGC	CTCCTTCAAG	GCCGTGTTCG	TGCCCGTGGC	CTACAGCCTC	180
	ATCTTCCTCC	TGGGCGTGAT	CGGCAACGTC	CTGGTGCTGG	TGATCCTGGA	GCGGCACCGG	240
	CAGACACGCA	GTTCCACGGA	GACCTTCCTG	TTCCACCTGG	CCGTGGCCGA	CCTCCTGCTG	300
15	GTCTTCATCT	TGCCCTTTGC	CGTGGCCGAG	GGCTCTGTGG	GCTGGGTCCT	GGGGACCTTC	360
	CTCTGCAAAA	CTGTGATTGC	CCTGCACAAA	GTCAACTTCT	ACTGCAGCAG	CCTGCTCCTG	420
	GCCTGCATCG	CCGTGGACCG	CTACCTGGCC	ATTGTCCACG	CCGTCCATGC	CTACCGCCAC	480
	CGCCGCCTCC	TCTCCATCCA	CATCACCTGT	GGGACCATCT	GGCTGGTGGG	CTTCCTCCTT	540
	GCCTTGCCAG	AGATTCTCTT	CGCCAAAGTC	AGCCAAGGCC	ATCACAACAA	CTCCCTGCCA	600
20	CGTTGCACCT	TCTCCCAAGA	GAACCAAGCA	GAAACGCATG	CCTGGTTCAC	CTCCCGATTC	660
	CTCTACCATG	TGGCGGGATT	CCTGCTGCCC	ATGCTGGTGA	TGGGCTGGTG	CTACGTGGGG	720
	GTAGTGCACA	GGTTGCGCCA	GGCCCAGCGG	CGCCCTCAGC	-GGCAGAAGGC	AGTCAGGGTG	780
	GCCATCCTGG	TGACAAGCAT	CTTCTTCCTC	TGCTGGTCAC	CCTACCACAT	CGTCATCTTC	840
	CTGGACACCC	TGGCGAGGCT	GAAGGCCGTG	GACAATACCT	GCAAGCTGAA	TGGCTCTCTC	900
25	CCCGTGGCCA	TCACCATGTG	TGAGTTCCTG	GGCCTGGCCC	ACTGCTGCCT	CAACCCCATG	960
	CTCTACACTT	TCGCCGGCGT	GAAGTTCCGC	AGTGACCTGT	CGCGGCTCCT	GACCAAGCTG	1020
	GGCTGTACCG	GCCCTGCCTC	CCTGTGCCAG	СТСТТСССТА	GCTGGCGCAG	GAGCAGTCTC	1080
	TCTGAGTCAG	AGAATGCCAC	CTCTCTCACC	ACGTTCTAG			1119

(67) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:

		(E) TY) SI	PE: RAND	I: 37 amin EDNE	o ac	id		٠,							
5	(i	.i) M	OLEC	ULE	TYPE	: pr	otei	n								
	(x	i) S	EQUE	NCE .	DESC	RIPT	'ION :	SEQ	ID	NO:6	6:					
	Met 1	. Asn	Tyr	Pro	Leu 5	Thr	Leu	Glu	Met	Asp 10	Leu	Glu	Asn	Leu	Glu 15	Asp
10	Leu	Phe	Trp	Glu 20	Leu	Asp	Arg	Leu	Asp 25	Asn	Tyr	Asn	Asp	Thr	Ser	Leu
	Val	Glu	Asn 35	His	Leu	Cys	Pro	Ala 40	Thr	Glu	Gly	Pro	Leu 45	Met	Ala	Ser
	Phe	Lys 50	Ala	Val	Phe	Val	Pro 55	Val	Ala	Tyr	Ser	Leu 60	Ile	Phe	Leu	Leu
15	Gly 65	Val	Ile	Gly	Asn	Val 70	Leu	Val	Leu	Val	Ile 75	Leu	Glu	Arg	His	Arg 80
	Gln	Thr	Arg	Ser	Ser 85	Thr	Glu	Thr	Phe	Leu 90	Phe	His	Leu	Ala	Val 95	Ala
20	Asp	Leu ,	Leu	Leu 100	Val	Phe	Ile	Leu	Pro 105	Phe	Ala	Val	Ala	Glu 110	Gly	Ser
	Val	Gly	Trp 115	Val	Leu	Gly	Thr	Phe 120	Leu	Cys	Lys	Thr	Val 125	Ile	Ala	Leu
	His	Lys 130	Val	Asn	Phe	Tyr	Cys 135	Ser	Ser	Leu	Leu	Leu 140	Ala	Cys	Ile	Ala
25	Val 145	Asp	Arg	Tyr	Leu	Ala 150	Ile	Val	His	Ala	Val 155	His	Ala	Tyr	Arg	His 160
		Arg			165	*				170					175	
30		Phe		180					185		•			190		
		His	195					200					205			
25		Ala 210					215					220				
35	225	Gly				230					235					240
	Val	Val	His	Ara	Leu	Ara	Gln	Ala	Gln	Δrα	Ara	Dro	C1 n	7 200	(1) n	T 110

BNSDOCID - WC 000010041

PCT/US99/23938

WO	ሰበ	177	170

						245					250					255		
	:	Ala	Val	Arg	Val 260	Ala	Ile	Leu	Val	Thr 265	Ser	Ile	Phe	Phe	Leu 270	Cys	Trp	
5	:	Ser	Pro	Tyr 275	His	Ile	Val	Ile	Phe 280	Leu	Asp	Thr	Leu	Ala 285	Arg	Leu	Lys	
	;	Ala	Val 290	Asp	Asn	Thr	Суѕ	Lys 295	Leu	Asn	Gly	Ser	Leu 300	Pro	Val	Ala	Ile	
		Thr 305	Met	Cys	Glu	Phe	Leu 310	Gly	Leu	Ala	His	Cys 315	Cys	Leu	Asn	Pro	Met 320	
10	:	Leu	Tyr	Thr	Phe	Ala 325	Gly	Val	Lys	Phe	Arg 330	Ser	Asp	Leu	Ser	Arg 335	Leu	
		Leu	Thr	Lys	Leu 340	Gly	Cys	Thr	Gly	Pro 345	Ala	Ser	Leu	Cys	Gln 350	Leu	Phe	
15	:	Pro	Ser	Trp 355	Arg	Arg	Ser	Ser	Leu 360	Ser	Glu	Ser	Glu	Asn 365	Ala	Thr	Ser	
	1	Leu	Thr 370	Thr	Phe											.,		;
	(68)	INFO	RMAT	rion	FOR	SEQ	ID 1	10:6	7:									
20		(i)	(A) (B) (C)	LEI TYI	CE CH NGTH: PE: 1 RANDI POLOG	: 30 nucle EDNES	base eic a SS: s	e par acid singl	irs									
		(ii	.) MC	DLECT	JLE :	TYPE	: DNZ	A (ge	enom:	ic)								
25		(xi	.) SI	EQUE	NCE I	DESC	RIPT	ION:	SEQ	ID 1	MO:61	7:			•			
	CAAAG	CTTG	SA AZ	AGCT	GCAC	GT(GCAG	AGAC										30
	(69)	INFC	RMAT	rion	FOR	SEQ	ID I	NO:68	3:									
30	·	(i)	(A) (B) (C)	LEI TYI	E CHANGTH: PE: 1 RANDI	: 30 nucle EDNES	base eic a SS: s	e pa: acid singl	irs			•	*					
		(ii	L) MC	OLEC	JLE :	TYPE	: DN	A (ge	enom	ic)						-		
		(xi	L) SI	EQUE	NCE I	DESCI	RIPT	ION:	SEQ	ID 1	NO : 68	3:						-
35	GCGGA	TCC	G AC	GTCA(CACC	TG	GCTG	GCC										30
	(70)	INFO	ORMA:	rion	FOR	SEO	ID I	NO:69	9:									

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1128 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5

- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

ATGGATGTGA CTTCCCAAGC CCGGGGCGTG GGCCTGGAGA TGTACCCAGG CACCGCGCAG 60 CCTGCGGCCC CCAACACCAC CTCCCCCGAG CTCAACCTGT CCCACCCGCT CCTGGGCACC 10 GCCCTGGCCA ATGGGACAGG TGAGCTCTCG GAGCACCAGC AGTACGTGAT CGGCCTGTTC CTCTCGTGCC TCTACACCAT CTTCCTCTTC CCCATCGGCT TTGTGGGCAA CATCCTGATC CTGGTGGTGA ACATCAGCTT CCGCGAGAAG ATGACCATCC CCGACCTGTA CTTCATCAAC 300 CTGGCGGTGG CGGACCTCAT CCTGGTGGCC GACTCCCTCA TTGAGGTGTT CAACCTGCAC GAGCGGTACT ACGACATCGC CGTCCTGTGC ACCTTCATGT CGCTCTTCCT GCAGGTCAAC 15 ATGTACAGCA GCGTCTTCTT CCTCACCTGG ATGAGCTTCG ACCGCTACAT CGCCCTGGCC 480 AGGGCCATGC GCTGCAGCCT GTTCCGCACC AAGCACCACG CCCGGCTGAG CTGTGGCCTC ATCTGGATGG CATCCGTGTC AGCCACGCTG GTGCCCTTCA CCGCCGTGCA CCTGCAGCAC 600 ACCGACGAGG CCTGCTTCTG TTTCGCGGAT GTCCGGGAGG TGCAGTGGCT CGAGGTCACG CTGGGCTTCA TCGTGCCCTT CGCCATCATC GGCCTGTGCT ACTCCCTCAT TGTCCGGGTG 720 20 CTGGTCAGGG CGCACCGGCA CCGTGGGCTG CGGCCCCGGC GGCAGAAGGC GCTCCGCATG 780 840 GTGCACCTCC TGCAGCGGAC GCAGCCTGGG GCCGCTCCCT GCAAGCAGTC TTTCCGCCAT 900 GCCCACCCC TCACGGGCCA CATTGTCAAC CTCACCGCCT TCTCCAACAG CTGCCTAAAC 960 CCCCTCATCT ACAGCTTTCT CGGGGAGACC TTCAGGGACA AGCTGAGGCT GTACATTGAG 1020 25 CAGAAAACAA ATTTGCCGGC CCTGAACCGC TTCTGTCACG CTGCCCTGAA GGCCGTCATT 1080 CCAGACAGCA CCGAGCAGTC GGATGTGAGG TTCAGCAGTG CCGTGTAG 1128

(71) INFORMATION FOR SEQ ID NO:70:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 375 amino acids
- (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant

		(ii) MC	LECU	LE I	YPE:	pro	tein		:.	*		•				
		(xi) SE	QUEN	ICE I	ESCF	IPTI	ON:	SEQ	ID N	10:70):					
		Met 1	Asp	Val	Thr	Ser 5	Gln	Ala	Arg	Gly	Val 10	Gly	Leu	Glu	Met	Tyr 15	Pro
5		Gly	Thr	Ala	Gln 20	Pro	Ala	Ala	Pro	Asn 25	Thr	Thr	Ser	Pro	Glu 30	Leu	Asn
		Leu	Ser	His 35	Pro	Leu	Leu	Gly	Thr 40	Ala	Leu	Ala	Asn	Gly 45	Thr	Gly	Glu
10		Leu	Ser 50	Glu	His	Gln	Gln	Tyr 55	Val	Ile	Gly	Leu	Phe 60	Leu	Ser	Cys	Leu
	•	Tyr 65	Thr	Ile	Phe	Leu	Phe 70	Pro	Ile	Gly	Phe	Val 75	Gly	Asn	Ile	Leu	Ile 80
		Leu	Val	Val	Asn	Ile 85	Ser	Phe	Arg	Glu	Lys 90	Met	Thr	Ile	Pro	Asp 95	Leu
15		Tyr	Phe	Ile	Asn 100	Leu	Ala	Val	Ala	Asp 105	Leu	Ile	Leu	Val	Ala 110	Asp	Ser
	-	Leu _.	Ile	Glu 115	Val	Phe	Asn	Leu	His 120	Glu	Arg	Tyr	Tyr	Asp 125	Ile	Ala	Val
20		Leu	Cys 130	Thr	Phe	Met	Ser	Leu 135	Phe	Leu	Gln	Val	Asn 140	Met	Tyr	Ser	Ser
		Val 145	Phe	Phe	Leu	Thr	Trp 150	Met	Ser	Phe	Asp	Arg 155	Tyr	Ile	Ala	Leu	Ala 160
. :	•	Arg	Ala	Met	Arg	Cys 165	Ser	Leu	Phe	Arg	Thr 170	Lys	His	His	Ala	Arg 175	Leu
25		Ser	Cys	Gly	Leu 180		Trp	Met	Ala	Ser 185		Ser	Ala	Thr	Leu 190		Pro
		Phe	Thr	Ala 195		His	Leu	Gln	His 200		Asp	Glu	Ala	Cys 205		Cys	Phe
30		Ala	Asp 210		Arg	Glu	Val	Gln 215		Leu	Glu	Val	Thr 220	Leu	Gly	Phe	Ile
		Val 225		Phe	Ala	Ile	230		Leu	Cys	Tyr	Ser 235		lle	Val	Arg	Val 240
		Leu	ı Val	. Arg	Ala	His 245		, His	Arg	g Gly	250		Pro	Arg	Arg	Gln 255	Lys
35		Ala	Lev	a Arg	Met 260		e Lev	a Ala	(Val	Val 265		val	. Phe	Phe	270	. Cys	Trp

		Leu	Pro	Glu 275	Asn	Val	Phe	Ile	Ser 280	Val	His	Leu	Leu	Gln 285	Arg	Thr	Gln	
		Pro	Gly 290	Ala	Ala	Pro	Cys	Lys 295	Gln	Ser	Phe	Arg	His 300	Ala	His	Pro	Leu	
. 5		Thr 305	Gly	His	Ile	Val	Asn 310	Leu	Thr	Ala	Phe	Ser 315	Asn	Ser	Сув	Leu	Asn 320	
		Pro	Leu	Ile	Tyr	Ser 325	Phe	Leu	Gly	Glu	Thr 330	Phe	Arg	Asp	Lys	Leu 335	Arg	
10		Leu	Tyr	Ile	Glu 340	Gln	Lys	Thr	Asn	Leu 345	Pro	Ala	Leu	Asn	Arg 350	Phe	Cys	
	4,	His	Ala	Ala 355	Leu	Lys	Ala	Val	Ile 360	Pro	Asp	Ser	Thr	Glu 365	Gln		Asp	
		Val	Arg 370	Phe	Ser	Ser	Ala	Val 375				ř				- 4 4		
15	(72)	INFO	ORMAT	rion	FOR	SEQ	ID 1	NO:71	L:									
20	. ** ***	(ii	(A) (B) (C) (D)	LEN TYI STI TOI	NGTH: PE: I RANDE POLOG JLE I	30 nucle EDNES EY: 1	base eic a ES: s linea	singl	rs .e.	.c)	JO:71	-1 ·	·					
	ACAG?	ATTO	CC TG	STGTO	GTTI	TAC	CCGC	CCAG									3	C
	(73)	INFO	RMAT	CION	FOR	SEQ	ID N	JO:72	::	٠			1		: •			
25		(i)	(A) (B) (C)	LEN TYP STR	IGTH: PE: r	30 ucle DNES	base ic a SS: s	ingl	rs			•		•	•			
30		(ii	.) MC	LECU	LE I	YPE:	DNA	(ge	nomi	.c)								
		(xi	.) SE	QUEN	ICE D	ESCR	IPTI	ON:	SEQ	ID N	0:72	:	٠					
	CTCGG	SATCO	'A GG	CAGA	AGAG	TCG	CCTA	TGG							•		3	0
	(74)	INFO	RMAT	'ION	FOR	SEQ	ID N	10:73	:									
35		(i)	(A) (B)	LEN TYP	GTH: E: n	113 ucle	7 ba	STIC se p cid	airs									

- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

<i>:</i> .	ATGGACCTGG	GGAAACCAAT	GAAAAGCGTG	CTGGTGGTGG	CTCTCCTTGT	CATTTTCCAG	60
5	GTATGCCTGT	GTCAAGATGA	GGTCACGGAC	GATTACATCG	GAGACAACAC	CACAGTGGAC	120
	TACACTTTGT.	TCGAGTCTTT	GTGCTCCAAG	AAGGACGTGC	GGAACTTTAA	AGCCTGGTTC	180
	CTCCCTATCA	TGTACTCCAT	CATTTGTTTC	GTGGGCCTAC	TGGGCAATGG	GCTGGTCGTG	240
	TTGACCTATA	TCTATTTCAA	GAGGCTCAAG	ACCATGACCG	ATACCTACCT	GCTCAACCTG	300
	GCGGTGGCAG	ACATCCTCTT	CCTCCTGACC	CTTCCCTTCT	GGGCCTACAG	CGCGGCCAAG	360
10	TCCTGGGTCT	TCGGTGTCCA	CTTTTGCAAG	CTCATCTTTG	CCATCTACAA	GATGAGCTTC	420
	TTCAGTGGCA	TGCTCCTACT	TCTTTGCATC	AGCATTGACC	GCTACGTGGC	CATCGTCCAG	480
	GCTGTCTCAG	CTCACCGCCA	CCGTGCCCGC	GTCCTTCTCA	TCAGCAAGCT	GTCCTGTGTG	540
	GGCATCTGGA	TACTAGCCAC	AGTGCTCTCC	ATCCCAGAGC	TCCTGTACAG	TGACCTCCAG	600
	AGGAGCAGCA	GTGAGCAAGC	GATGCGATGC	TCTCTCATCA	CAGAGCATGT	GGAGGCCTTT	660
15	ATCACCATCC	AGGTGGCCCA	GATGGTGATC	GGCTTTCTGG	TCCCCCTGCT	GGCCATGAGC	720
	TTCTGTTACC	TTGTCATCAT	CCGCACCCTG	CTCCAGGCAC	GCAACTTTGA	GCGCAACAAG	780
	GCCATCAAGG	TGATCATCGC	TGTGGTCGTG	GTCTTCATAG	TCTTCCAGCT	GCCCTACAAT	840
	GGGGTGGTCC	TGGCCCAGAC	GGTGGCCAAC	TTCAACATCA	CCAGTAGCAC	CTGTGAGCTC	900
	AGTAAGCAAC	TCAACATCGC	CTACGACGTC	ACCTACAGCC	TGGCCTGCGT	CCGCTGCTGC	960
20	GTCAACCCTT	TCTTGTACGC	CTTCATCGGC	GTCAAGTTCC	GCAACGATCT	CTTCAAGCTC	1020
	TTCAAGGACC	TGGGCTGCCT	CAGCCAGGAG	CAGCTCCGGC	AGTGGTCTTC	CTGTCGGCAC	1080
	ATCCGGCGCT	CCTCCATGAG	TGTGGAGGCC	GAGACCACCA	CCACCTTCTC	CCCATAG	1137

(75) INFORMATION FOR SEQ ID NO:74:

- (i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 378 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein
- 30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

	Met 1	Asp	Leu	Gly	Lys 5	Pro	Met	Lys	Ser	Val	Leu	ı Val	Val	. Ala	Leu 15	Leu	
	Val	Ile	Phe	Gln 20	Val	Cys	Leu	Cys	Cln 25	. Asp	Glu	Val	Thr	Asp 30	Asp	Tyr	
5	Ile	Gly	Asp 35	Asn	Thr	Thr	Val	Asp 40	Tyr	Thr	Leu	Phe	Glu 45	Ser	Leu	Cys	
	Ser	Lys 50	Lys	Asp	Val	Arg	Asn 55	Phe	Lys	Ala	Trp	Phe 60	Leu	Pro	Ile	Met	
10	Tyr 65	Ser	Ile	Ile	Cys	Phe 70	Val	Gly	Leu	Leu	Gly 75	Asn	Gly	Leu	Val	Val 80	
	Leu	Thr	Tyr	Ile	Tyr 85	Phe	Lys	Arg	Leu	Lys 90	Thr	Met	Thr	Asp	Thr 95	Tyr	
	Leu	Leu	Asn	Leu 100	Ala	Val	Ala	Asp	Ile 105	Leu	Phe	Leu	Leu	Thr 110	Leu	Pro	
15	Phe	Trp	Ala 115	Tyr	Ser	Ala	Ala	Lys 120	Ser	Trp	Val	Phe	Gly 125	Val	His	Phe	
	Cys	Lys 130	Leu	Ile	Phe		Ile 135	Tyr	Lys	Met	Ser	Phe 140	Phe	Ser	Gly	Met	
20	Leu 145	Leu	Leu	Leu	Cys	Ile 150	Ser	Ile	Asp	Arg	Tyr 155	Val	Ala	Ile	Val	Gln 160	
	Ala	Val	Ser	Ala	His 165	Arg	His	Arg	Ala	Arg 170	Val	Leu	Leu	Ile	Ser 175	Lys	
	Leu	Ser	Cys	Val 180	Gly	Ile	Trp	Ile	Leu 185	Ala	Thr	Val	Leu	Ser 190	Ile	Pro	
25	Glu	Leu	Leu 195	Tyr	Ser	Asp	Leu	Gln 200	Arg	Ser	Ser	Ser	Glu 205	Gln	Ala	Met	
	Arg	Cys 210	Ser	Leu	Ile	Thr	Glu 215	His	Val	Glu	Ala	Phe 220	Ile	Thr	Ile	Gln	
30	Val 225	Ala	Gln	Met	Val	Ile 230	Gly	Phe	Leu	Val	Pro 235	Leu	Leu	Ala	Met	Ser 240	
	Phe	Cys	Tyr	Leu	Val 245	Ile	Ile	Arg	Thr	Leu 250	Leu	Gln	Ala	Arg	Asn 255	Phe	
	Glu	Arg	Asn	Lys 260	Ala	Ile	Lys	Val	Ile 265	Ile	Ala	Val	Val	Val 270	Val	Phe	
35	Ile	Val	Phe 275	Gln	Leu	Pro	Tyr	Asn 280	Gly	Val	Val	Leu	Ala	Gln	Thr	Val	

	Ala Asn Phe Asn Ile Thr Ser Ser Thr Cys Glu Leu Ser Lys Gln Leu 290 295 300	Į.
	Asn Ile Ala Tyr Asp Val Thr Tyr Ser Leu Ala Cys Val Arg Cys Cys 305 310 315 320	
5	Val Asn Pro Phe Leu Tyr Ala Phe Ile Gly Val Lys Phe Arg Asn Asp 325 330 335	,
	Leu Phe Lys Leu Phe Lys Asp Leu Gly Cys Leu Ser Gln Glu Gln Leu 340 345 350	
10	Arg Gln Trp Ser Ser Cys Arg His Ile Arg Arg Ser Ser Met Ser Val	
	Glu Ala Glu Thr Thr Thr Phe Ser Pro	
	(76) INFORMATION FOR SEQ ID NO:75:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 32 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	٠.
	(ii) MOLECULE TYPE: DNA (genomic)	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:	v;
	CTGGAATTCA CCTGGACCAC CACCAATGGA TA	32
	(77) INFORMATION FOR SEQ ID NO:76:	
25	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:	
30	CTCGGATCCT GCAAAGTTTG TCATACAGTT	30
	(78) INFORMATION FOR SEQ ID NO:77:	
35	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1085 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

	ATGGATATAC	AAATGGCAAA	CAATTTTACT	CCGCCCTCTG	CAACTCCTC	GGGAAATGAC	60
	TGTGACCTCT	' ATGCACATCA	CAGCACGGCC	AGGATAGTAA	TGCCTCTGCA	TTACAGCCTC	120
	GTCTTCATCA	TTGGGCTCGT	GGGAAACTTA	CTAGCCTTGG	TCGTCATTGT	'TCAAAACAGG	180
5	AAAAAAATCA	ACTCTACCAC	CCTCTATTCA	ACAAATTTGG	TGATTTCTGA	TATACTTTTT	240
	ACCACGGCTT	TGCCTACACG	AATAGCCTAC	TATGCAATGG	GCTTTGACTG	GAGAATCGGA	300
	GATGCCTTGT	GTAGGATAAC	TGCGCTAGTG	TTTTACATCA	ACACATATGO	AGGTGTGAAC	360
	TTTATGACCT	GCCTGAGTAT	TGACCGCTTC	ATTGCTGTGG	TGCACCCTCT	ACGCTACAAC	420
	AAGATAAAA	GGATTGAACA	TGCAAAAGGC	GTGTGCATAT	TTGTCTGGAT	TCTAGTATTT	480
10	GCTCAGACAC	TCCCACTCCT	CATCAACCCT	ATGTCAAAGC	AGGAGGCTGA	AAGGATTACA	540
	TGCATGGAGT	ATCCAAACTT	TGAAGAAACT	AAATCTCTTC	CCTGGATTCT	GCTTGGGGCA	600
	TGTTTCATAG	GATATGTACT	TCCACTTATA	ATCATTCTCA	TCTGCTATTC	TCAGATCTGC	660
	TGCAAACTCT	TCAGAACTGC	CAAACAAAAC	CCACTCACTG	AGAAATCTGG	ТСТАЛАСАЛА	720
	AAGGCTCTCA	ACACAATTAT	TCTTATTATT	GTTGTGTTTG	TTCTCTGTTT	CACACCTTAC	780
15	CATGTTGCAA	TTATTCAACA	TATGATTAAG	AAGCTTCGTT	TCTCTAATTT	CCTGGAATGT	840
	AGCCAAAGAC	ATTCGTTCCA	GATTTCTCTG	CACTTTACAG	TATGCCTGAT	GAACTTCAAT	900
	TGCTGCATGG	ACCCTTTTAT	CTACTTCTTT	GCATGTAAAG	GGTATAAGAG	AAAGGTTATG	960
	AGGATGCTGA	AACGGCAAGT	CAGTGTATCG	ATTTCTAGTG	CTGTGAAGTC	AGCCCCTGAA	1020
	GAAAATTCAC	GTGAAATGAC	AGAAACGCAG	ATGATGATAC	ATTCCAAGTC	TTCAAATGGA	1080
20	AAGTGA			•		e a	1086

(79) INFORMATION FOR SEQ ID NO:78:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 361 amino acids
 - (B) TYPE: amino acid
- 25 (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

Met Asp Ile Gln Met Ala Asn Asn Phe Thr Pro Pro Ser Ala Thr Pro 1 5 10 15

	Gln	Gly	Asn	Asp 20	Cys	Asp	Leu	Tyr	Ala 25	His	His	Ser	Thr	Ala 30	Arg	Ile
	Val	Met	Pro 35	Leu	His	Tyr	Ser	Leu 40	Val	Phe	Ile	Ile	Gly 45	Leu	Val	Gly
5	Asn	Leu 50	Leu	Ala	Leu	Val	Val 55	Ile	Val	Gln	Asn	Arg 60	Lys	Lys	Ile	Asn
	Ser 65	Thr	Thr	Leu	Ťyr	Ser 70	Thr	Asn	Leu	Val	Ile 75	Ser	Asp	Ile	Leu	Phe 80
10	Thr	Thr	Ala	Leu	Pro 85	Thr	Arg	Ile	Ala	Tyr 90	Tyr	Ala	Met	Gly	Phe 95	Asp
	Trp	Arg	Ile	Gly 100	Asp	Ala	Leu	Cys	Arg 105	Ile	Thr	Ala	Leu	Val 110	Phe	Tyr
	Ile	Asn	Thr 115	Tyr	Ala	Gly	Val	Asn 120	Phe	Met	Thr	Cys	Leu 125	Ser	Ile	Asp
15	Arg	Phe 130	Ile	Ala	Val	Val	His 135	Pro	Leu	Arg	Tyr	Asn 140	Lys	Ile	Lys	Arg
	Ile 145	Glu	His	Ala	Lys	Gly 150	Val	Cys	Ile	Phe	Val 155	Trp	Ile	Leu	Val	Phe 160
20	Ala	Gln	Thr	Leu	Pro 165	Leu	Leu	Ile	Asn	Pro 170	Met	Ser	Lys	Gln	Glu 175	Ala
	Glu	Arg	Ile	Thr 180	Cys	Met	Glu	Tyr	Pro 185	Asn	Phe	Glu	Glu	Thr 190	Lys	Ser
	Leu	Pro	Trp 195	Ile	Leu	Leu	Gly	Ala 200	Cys	Phe	Ile	Gly	Tyr 205	Val	Leu	Pro
25	Leu	Ile 210	•	Ile	Leu		Cys 215	Tyr	Ser	Gln	Ile	Cys 220	Cys	Lys	Leu	Phe
	Arg 225	Thr	Ala	Lys	Gln	Asn 230	Pro	Leu	Thr	Glu	Lys 235	Ser	Gly	Val	Asn	Lys 240
30	Lys	Ala	Leu	Asn	Thr 245	Ile	Ile	Leu	Ile	11e 250	Val	Val	Phe	Val	Leu 255	Cys
	Phe	Thr	Pro	Tyr 260		Val	Ala	Ile	Ile 265		His	Met	Ile	Lys 270	Lys	Leu
	Arg	Phe	Ser 275		Phe	Leu	Glu	Cys 280		Gln	Arg	His	Ser 285		Gln	Ile
35	Ser	Leu 290		Phe	Thr	Val	Cys 295		Met	Asn	Phe	Asn 300		Cys	Met	Ası
	Pro	Phe	Ile	Tyr	Phe	Phe	Ala	Cys	Lys	Gly	Tyr	Lys	Arg	Lys	Val	Me

	305 310 315 320	
	Arg Met Leu Lys Arg Gln Val Ser Val Ser Ile Ser Ser Ala Val Lys 325 330 335	
5	Ser Ala Pro Glu Glu Asn Ser Arg Glu Met Thr Glu Thr Gln Met Met 340 345 350	
	Ile His Ser Lys Ser Ser Asn Gly Lys 355 360	
	(80) INFORMATION FOR SEQ ID NO:79:	
10	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
15	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:	
	CTGGAATTCT CCTGCTCATC CAGCCATGCG G	31
	(81) INFORMATION FOR SEQ ID NO:80:	
20	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:	
25	CCTGGATCCC CACCCCTACT GGGGCCTCAG	30
	(82) INFORMATION FOR SEQ ID NO:81:	
30	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1446 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:	
	ATGCGGTGGC TGTGGCCCCT GGCTGTCTCT CTTGCTGTGA TTTTGGCTGT GGGGCTAAGC	60
35	AGGGTCTCTG GGGGTGCCCC CCTGCACCTG GGCAGGCACA GAGCCGAGAC CCAGGAGCAG	120

BRICHAND SWID DUSSESSEE IN

	CAGAGCCGAT	CCAAGAGGGG	CACCGAGGAT	GAGGAGGCCA	AGGGCGTGCA	GCAGTATGTG	180
	CCTGAGGAGT	GGGCGGAGTA	CCCCGGCCC	ATTCACCCTG	CTGGCCTGCA	GCCAACCAAG	240
	CCCTTGGTGG	CCACCAGCCC	TAACCCCGAC	AAGGATGGGG	GCACCCCAGA	CAGTGGGCAG	300
	GAACTGAGGG	GCAATCTGAC	AGGGGCACCA	GGGCAGAGGC	TACAGATCCA	GAACCCCCTG	360
5	TATCCGGTGA	CCGAGAGCTC	CTACAGTGCC	TATGCCATCA	TGCTTCTGGC	GCTGGTGGTG	420
	TTTGCGGTGG	GCATTGTGGG	CAACCTGTCG	GTCATGTGCA	TCGTGTGGCA	CAGCTACTAC	480
	CTGAAGAGCG	CCTGGAACTC	CATCCTTGCC	AGCCTGGCCC	TCTGGGATTT	TCTGGTCCTC	540
	TTTTTCTGCC	TCCCTATTGT	CATCTTCAAC	GAGATCACCA	AGCAGAGGCT	ACTGGGTGAC	600
	GTTTCTTGTC	GTGCCGTGCC	CTTCATGGAG	GTCTCCTCTC	TGGGAGTCAC	GACTTTCAGC	660
0	CTCTGTGCCC	TGGGCATTGA	CCGCTTCCAC	GTGGCCACCA	GCACCCTGCC	CAAGGTGAGG	720
	CCCATCGAGC	GGTGCCAATC	CATCCTGGCC	AAGTTGGCTG	TCATCTGGGT	GGGCTCCATG	780
	ACGCTGGCTG	TGCCTGAGCT	CCTGCTGTGG	CAGCTGGCAC	AGGAGCCTGC	CCCCACCATG	840
	GGCACCCTGG	ACTCATGCAT	CATGAAACCC	TCAGCCAGCC	TGCCCGAGTC	CCTGTATTCA	900
	CTGGTGATGA	CCTACCAGAA	CGCCCGCATG	TGGTGGTACT	TTGGCTGCTA	CTTCTGCCTG	960
5	CCCATCCTCT	TCACAGTCAC	CTGCCAGCTG	GTGACATGGC	GGGTGCGAGG	CCCTCCAGGG	1020
	AGGAAGTCAG	AGTGCAGGGC	CAGCAAGCAC	GAGCAGTGTG	AGAGCCAGCT	CAACAGCACC	1080
	GTGGTGGGCC	TGACCGTGGT	CTACGCCTTC	TGCACCCTCC	CAGAGAACGT	CTGCAACATC	1140
	GTGGTGGCCT	ACCTCTCCAC	CGAGCTGACC	CGCCAGACCC	TGGACCTCCT	GGGCCTCATC	1200
	AACCAGTTCT	CCACCTTCTT	CAAGGCCCC	ATCACCCCAG	TGCTGCTCCT	TTGCATCTGC	1260
20	AGGCCGCTGG	GCCAGGCCTT	CCTGGACTGC	TGCTGCTGCT	GCTGCTGTGA	GGAGTGCGGC	1320
	GGGGCTTCGG	AGGCCTCTGC	TGCCAATGGG	TCGGACAACA	AGCTCAAGAC	CGAGGTGTCC	1380
	TCTTCCATCT	ACTTCCACAA	GCCCAGGGAG	TCACCCCAC	TCCTGCCCCT	GGGCACACCT	1440
	TGCTGA						1446

(83) INFORMATION FOR SEQ ID NO:82:

25 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 481 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- 30 (ii) MOLECULE TYPE: protein

	()	i) s	EQUE	NCE	DESC	RIPI	'ION:	SEQ	ID	NO : 8	2:					
	Met 1	Arg	Trp	Leu	Trp 5	Pro	Leu	Ala	Val	Ser	Leu	Ala	Val	Ile	Leu 15	Ala
5	Val	Gly	Leu	Ser 20	Arg	Val	Ser	Gly	Gly 25	Ala	Pro	Leu	His	Leu 30	Gly	Arg
	His	Arg	Ala 35	Glu	Thr	Gln	Glu	Gln 40	Gln	Ser	Arg	Ser	Lys 45	Arg	Gly	Thr
	Glu	Asp 50	Glu	Glu	Ala	Lys	Gly 55	Val	Gln	Gln	Tyr	Val 60	Pro	Glu	Glu	Trp
10	Ala 65	Glu	Tyr	Pro	Arg	Pro 70	Ile	His	Pro	Ala	Gly 75	Leu	Gln	Pro	Thr	Lys 80
	Pro	Leu	Val.	Ala	Thr 85	Ser	Pro	Asn	Pro	Asp 90	Lys	Asp	Gly	Gly	Thr 95	Pro
15	Asp	Ser	Gly	Gln 100	Glu	Leu	Arg	Gly	Asn 105	Leu	Thr	Gly	Ala	Pro	Gly	Gln
,	Arg	Leu	Gln 115	Ile	Gln	Asn	Pro	Leu 120	Tyr	Pro	Val	Thr	Glu 125	Ser	Ser	Tyr
	Ser	Ala 130	Tyr	Ala	Ile	Met	Leu 135	Leu	Ala	Leu	Val	Val 140	Phe	Ala	Val	Gly
20	Ile 145	Val	Gly	Asn	Leu	Ser 150	Val	Met	Cys	Ile	Val 155	Trp	His	Ser	Tyr	Tyr 160
	Leu	Lys	Ser	Ala	Trp 165	Asn	Ser	Ile	Leu	Ala 170	Ser	Leu	Ala	Leu	Trp 175	Asp
25	Phe	Leu	Val	Leu 180	Phe	Phe	Суя	Leu	Pro 185	Ile	Val	Ile	Phe	Asn 190	Glu	Ile
	Thr	Lys	Gln	Arg	Leu	Leu	Glv	azA	Val	Ser	Cvs	Ara	Δla	17=1	Dro	Dho

Thr Lys Gln Arg Leu Leu Gly Asp Val Ser Cys Arg Ala Val Pro Phe Met Glu Val Ser Ser Leu Gly Val Thr Thr Phe Ser Leu Cys Ala Leu Gly Ile Asp Arg Phe His Val Ala Thr Ser Thr Leu Pro Lys Val Arg Pro Ile Glu Arg Cys Gln Ser Ile Leu Ala Lys Leu Ala Val Ile Trp Val Gly Ser Met Thr Leu Ala Val Pro Glu Leu Leu Leu Trp Gln Leu Ala Gln Glu Pro Ala Pro Thr Met Gly Thr Leu Asp Ser Cys Ile Met

BNSDOCID - CWD 000010941 14-

				275					280					285			
		Lys	Pro 290	Ser	Ala	Ser 	Leu	Pro 295	Glu	Ser	Leu	Tyr	Ser 300	Leu	Val	Met	Thr
5		Tyr 305	Gln	Asn	Ala	Arg	Met 310	Trp	Trp	Tyr	Phe	Gly 315	Cys	Tyr	Phe	Cys	Leu 320
		Pro	Ile	Leu	Phe	Thr 325	Val	Thr	Cys	Gln	Leu 330	Val	Thr	Trp	Arg	Val 335	Arg
		Gly	Pro	Pro	Gly 340	Arg	Lys	Ser	Glu	Cys 345	Arg	Ala	Ser	Lys	His 350	Glu	Gln
10		Cys	Glu	Ser 355	Gln	Leu	Asn	Ser	Thr 360	Val	Val	Gly	Leu	Thr 365	Val	Val	Tyr
		Ala	Phe 370	Cys	Thr	Leu	Pro	Glu 375	Asn	Val	Cys	Asn	Ile 380	Val	Val	Ala	Tyr
15		Leu 385	Ser	Thr	Glu	Leu	Thr 390	Arg	Gln	Thr	Leu	Asp 395	Leu	Leu	Gly	Leu	Ile 400
		Asn	Gln	Phe	Ser	Thr 405	Phe	Phe	Lys	Gly	Ala 410	Ile	Thr	Pro	Val	Leu 415	Leu
	÷	Leu	Cys	Ile	Cys 420	Arg	Pro	Leu	Gly	Gln 425	Ala	Phe	Leu	Asp	Cys 430	Cys	Cys
20		Cys	-	Cys 435	Cys	Glu	Glu	Cys	Gly 440	Gly	Ala	Ser	Glu	Ala 445	Ser	Ala	Ala
		Asn	Gly 450	Ser	Asp	Asn	Lys	Leu 455	Lys	Thr	Glu	Val	Ser 460	Ser	Ser	Ile	Tyr
25	٠	Phe 465	His	Lys	Pro	Arg	Glu 470	Ser	Pro	Pro	Leu	Leu 475	Pro	Leu	Gly	Thr	Pro 480
		Cys											:				
	(84)	INFO	ORMA!	rion	FOR	SEQ	ID 1		3:			·					
30		(i)	(A) (B) (C)) LEI) TYI) STI	NGTH PE: 1 RANDI	: 22 nucle EDNE	base eic a	ISTIC e pa: acid sing:	CS: irs								
		(i:	i) M	OLEC	ULE :	TYPE	: DN	A (g	enom:	ic) ´							
35		(x:	i) SI	EQUEI	NCE I	DESC	RIPT	ION:	SEQ	ID 1	NO : 8	3:					

	(85) INFORMATION FOR SEQ ID NO:84:	
5	(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 22 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:	
	TCATGTATTA ATACTAGATT CT	22
10	(86) INFORMATION FOR SEQ ID NO:85:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 38 base pairs	
	(B) TYPE: nucleic acid (C) STRANDEDNESS: single	
15	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:	
	TACCATGTGG AACGCGACGC CCAGCGAAGA GCCGGGGT	38
	(87) INFORMATION FOR SEQ ID NO:86:	
20	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
25	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:	
	CGGAATTCAT GTATTAATAC TAGATTCTGT CCAGGCCCG	39
	(88) INFORMATION FOR SEQ ID NO:87:	
20	(i) SEQUENCE CHARACTERISTICS:	
30	(A) LENGTH: 1101 base pairs (B) TYPE: nucleic acid	
•	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:	

BNSDOOID -WIO 000010041 14.

	ATGTGGAACG	CGACGCCCAG	CGAAGAGCCG	GGGTTCAACC	TCACACTGGC	CGACCTGGAC	6
	TGGGATGCTT	CCCCCGGCAA	CGACTCGCTG	GGCGACGAGC	TGCTGCAGCT	CTTCCCCGCG	120
	CCGCTGCTGG	CGGGCGTCAC	AGCCACCTGC	GTGGCACTCT	TCGTGGTGGG	TATCGCTGGC	180
	AACCTGCTCA	CCATGCTGGT	GGTGTCGCGC	TTCCGCGAGC	TGCGCACCAC	CACCAACCTC	240
5	TACCTGTCCA	GCATGGCCTT	CTCCGATCTG	CTCATCTTCC	TCTGCATGCC	CCTGGACCTC	300
	GTTCGCCTCT	GGCAGTACCG	GCCCTGGAAC	TTCGGCGACC	TCCTCTGCAA	ACTCTTCCAA	360
	TTCGTCAGTG	AGAGCTGCAC	CTACGCCACG	GTGCTCACCA	TCACAGCGCT	GAGCGTCGAG	420
	CGCTACTTCG	CCATCTGCTT	CCCACTCCGG	GCCAAGGTGG	TGGTCACCAA	GGGGCGGGTG	480
	AAGCTGGTCA	TCTTCGTCAT	CTGGGCCGTG	GCCTTCTGCA	GCGCCGGGCC	CATCTTCGTG	540
10	CTAGTCGGGG	TGGAGCACGA	GAACGGCACC	GACCCTTGGG	ACACCAACGA	GTGCCGCCCC	600
	ACCGAGTTTG	CGGTGCGCTC	TGGACTGCTC	ACGGTCATGG	TGTGGGTGTC	CAGCATCTTC	660
	TTCTTCCTTC	CTGTCTTCTG	TCTCACGGTC	CTCTACAGTC	TCATCGGCAG	GAAGCTGTGG	720
	CGGAGGAGGC	GCGGCGATGC	TGTCGTGGGT	GCCTCGCTCA	GGGACCAGAA	CCACAAGCAA	780
	ACCGTGAAAA	TGCTGGCTGT	AGTGGTGTTT	GCCTTCATCC	TCTGCTGGCT	CCCCTTCCAC	840
15	GTAGGGCGAT	ATTTATTTTC	CAAATCCTTT	GAGCCTGGCT	CCTTGGAGAT	TGCTCAGATC	900
	AGCCAGTACT	GCAACCTCGT	GTCCTTTGTC	CTCTTCTACC	TCAGTGCTGC	CATCAACCCC	960
	ATTCTGTACA	ACATCATGTC	CAAGAAGTAC	CGGGTGGCAG	TGTTCAGACT	TCTGGGATTC	1020
	GAACCCTTCT	CCCAGAGAAA	GCTCTCCACT	CTGAAAGATG	AAAGTTCTCG	GGCCTGGACA	1080
	GAATCTAGTA	TTAATACATG	А	,	•	•	1101

 $20\,$ (89) Information for SEQ ID NO:88:

25

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 366 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

Met Trp Asn Ala Thr Pro Ser Glu Glu Pro Gly Phe Asn Leu Thr Leu 1 5 10 15

Ala Asp Leu Asp Trp Asp Ala Ser Pro Gly Asn Asp Ser Leu Gly Asp 20 25 30

	•	Glu	Leu	Leu 35	Gln	Leu	Phe	Pro	Ala 40	Pro	Leu	Leu	Ala	Gly 45	Val	Thr	Ala
		Thr	Cys 50	Val	Ala	Leu	Phe	Val 55	Val	Gly	Ile	Ala	Gly 60	Asn	Leu	Leu	Th
5		Met 65	Leu	Val	Val	Ser	Arg 70	Phe	Arg	Glu	Leu	Arg 75	Thr	Thr	Thr	Asn	Le:
		Tyr	Leu	Ser	Ser	Met 85	Ala	Phe	Ser	Asp	Leu 90	Leu	Ile	Phe	Leu	Cys 95	Met
10		Pro	Leu	Asp	Leu 100	Val	Arg	Leu	Trp	Gln 105	Tyr	Arg	Pro	Trp	Asn 110	Phe	Gl
		Asp	Leu	Leu 115	Cys	Lys	Leu	Phe	Gln 120	Phe	Val	Ser	Glu	Ser 125	Cys	Thr	Туз
		Ala	Thr 130	Val	Leu	Thr	Ile	Thr 135	Ala	Leu	Ser	Val	Glu 140	Arg	Tyr	Phe	Ala
15		Ile 145	Cys	Phe	Pro	Leu	Arg 150	Ala	Lys	Val	Val	Val 155	Thr	Lys	Gly	Arg	Va]
	••	Lys	Leu	Val	Ile	Phe 165	Val	Ile	Trp	Ala	Val 170	Ala	Phe	Cys	Ser	Ala 175	Gly
20		Pro	Ile	Phe	Val 180	Leu	Val	Gly	Val	Glu 185	His	Glu	Asn	Gly	Thr 190	Asp	Pro
		Trp	Asp	Thr 195	Asn	Glu	Cys	Arg	Pro 200	Thr	Glu	Phe	Ala	Val 205	Arg	Ser	Gl
		Leu	Leu 210	Thr	Val	Met	Val	Trp 215	Val	Ser	Ser	Ile	Phe 220		Phe	Leu	Pro
25		Val 225	Phe	Cys	Leu	Thr	Val 230	Leu	Tyr	Ser	Leu	Ile 235	Gly	Arg	Lys	Leu	Trp 240
		Arg	Arg	Arg	Arg	Gly 245	Asp	Ala	Val	Val	Gly 250	Ala	Ser	Leu	Arg	Asp 255	Glr
30		Asn	His	Lys	Gln 260	Thr	Val	Lys	Met	Leu 265	Ala	Val	Val	Val	Phe 270	Ala	Phe
				275					280					285		Ser	
3.5		-	290					295					300			Tyr	-
35		305					310					315				Asn	320
		TIE	1.611	Tvr	Asn	IIe	Met	Ser	LVC	1.376	Tur	Ara	1721	Δla	17 = 1	Dho	Arc

PCT/US99/23938

v	À	ΛΛ	/22	17	O

	325 330 335	
	Leu Leu Gly Phe Glu Pro Phe Ser Gln Arg Lys Leu Ser Thr Leu Ly 340 345 350	rs
5.	Asp Glu Ser Ser Arg Ala Trp Thr Glu Ser Ser Ile Asn Thr 355 360 365	
	(90) INFORMATION FOR SEQ ID NO:89:	
10	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:	
	GCAAGCTTGT GCCCTCACCA AGCCATGCGA GCC	33
15	(91) INFORMATION FOR SEQ ID NO:90:	
	(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 30 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single	£.
20	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:	
	CGGAATTCAG CAATGAGTTC CGACAGAAGC	30
	(92) INFORMATION FOR SEQ ID NO:91:	
25	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1842 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
30	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:	
	ATGCGAGCCC CGGGCGCGCT TCTCGCCCGC ATGTCGCGGC TACTGCTTCT GCTACTGCTC	60
	AAGGTGTCTG CCTCTCTGC CCTCGGGGTC GCCCCTGCGT CCAGAAACGA AACTTGTCTG	120
	GGGGAGAGCT GTGCACCTAC AGTGATCCAG CGCCGCGGCA GGGACGCCTG GGGACCGGGA	180
35	AATTCTGCAA GAGACGTTCT GCGAGCCCGA GCACCCAGGG AGGAGCAGGG GGCAGCGTTT	240

CTTGCGGGAC CCTCCTGGGA CCTGCCGGCG GCCCCGGGCC GTGACCCGGC TGCAGGCAGA 300 GGGGCGGAGG CGTCGGCAGC CGGACCCCG GGACCTCCAA CCAGGCCACC TGGCCCCTGG AGGTGGAAAG GTGCTCGGGG TCAGGAGCCT TCTGAAACTT TGGGGAGAGG GAACCCCACG 420 GCCCTCCAGC TCTTCCTTCA GATCTCAGAG GAGGAAGAGA AGGGTCCCAG AGGCGCTGGC 480 ATTTCCGGGC GTAGCCAGGA GCAGAGTGTG AAGACAGTCC CCGGAGCCAG CGATCTTTTT 540 TACTGGCCAA GGAGAGCCGG GAAACTCCAG GGTTCCCACC ACAAGCCCCT GTCCAAGACG 600 GCCAATGGAC TGGCGGGCA CGAAGGGTGG ACAATTGCAC TCCCGGGCCG GGCGCTGGCC CAGAATGGAT CCTTGGGTGA AGGAATCCAT GAGCCTGGGG GTCCCCGCCG GGGAAACAGC 720 ACGAACCGGC GTGTGAGACT GAAGAACCCC TTCTACCCGC TGACCCAGGA GTCCTATGGA 780 10 GCCTACGCGG TCATGTGTCT GTCCGTGGTG ATCTTCGGGA CCGGCATCAT TGGCAACCTG 840 GCGGTGATGA GCATCGTGTG CCACAACTAC TACATGCGGA GCATCTCCAA CTCCCTCTTG 900 GCCAACCTGG CCTTCTGGGA CTTTCTCATC ATCTTCTTCT GCCTTCCGCT GGTCATCTTC CACGAGCTGA CCAAGAAGTG GCTGCTGGAG GACTTCTCCT GCAAGATCGT GCCCTATATA 1020 GAGGTCGCTT CTCTGGGAGT CACCACTTTC ACCTTATGTG CTCTGTGCAT AGACCGCTTC 1080 15 CGTGCTGCCA CCAACGTACA GATGTACTAC GAAATGATCG AAAACTGTTC CTCAACAACT GCCAAACTTG CTGTTATATG GGTGGGAGCT CTATTGTTAG CACTTCCAGA AGTTGTTCTC 1200 CGCCAGCTGA GCAAGGAGGA TTTGGGGTTT AGTGGCCGAG CTCCGGCAGA AAGGTGCATT 1260 ATTAAGATCT CTCCTGATTT ACCAGACAC ATCTATGTTC TAGCCCTCAC CTACGACAGT 1320 GCGAGACTGT GGTGGTATTT TGGCTGTTAC TTTTGTTTGC CCACGCTTTT CACCATCACC 1380 TGCTCTCTAG TGACTGCGAG GAAAATCCGC AAAGCAGAGA AAGCCTGTAC CCGAGGGAAT 1440 20 AAACGGCAGA TTCAACTAGA GAGTCAGATG AACTGTACAG TAGTGGCACT GACCATTTTA 1500 TATGGATTTT GCATTATTCC TGAAAATATC TGCAACATTG TTACTGCCTA CATGGCTACA 1560 GGGGTTTCAC AGCAGACAAT GGACCTCCTT AATATCATCA GCCAGTTCCT TTTGTTCTTT AAGTCCTGTG TCACCCCAGT CCTCCTTTTC TGTCTCTGCA AACCCTTCAG TCGGGCCTTC 1680 ATGGAGTGCT GCTGCTGTTG CTGTGAGGAA TGCATTCAGA AGTCTTCAAC GGTGACCAGT 1740 GATGACAATG ACAACGAGTA CACCACGGAA CTCGAACTCT CGCCTTTCAG TACCATACGC 1800 CGTGAAATGT CCACTTTTGC TTCTGTCGGA ACTCATTGCT GA 1842

5		(A) (B) (C) (D)	LEN TYP STR TOP	IGTH: PE: a RANDE PCLOG	613 minc DNES	ami aci S: ot r	STIC no a .d relev	cids vant	· :		-					
	(xi	.) SE	EQUEN	ICE I	ESCF	RIPTI	ON:	SEQ	ID N	10:92	2:					
	Met 1	Arg	Ala	Pro	Gly 5	Ala	Leu	Leu	Ala	Arg 10	Met	Ser	Arg	Leu	Leu 15	Leu
10	Leu	Leu	Leu	Leu 20	Lys	Val	Ser	Ala	Ser 25	Ser	Ala	Leu	Gly	Val 30	Ala	Pro
	Ala	Ser	Arg 35	Asn	Glu	Thr	Cys	Leu 40	Gly	Glu	Ser	Cys	Ala 45	Pro	Thr	Val
15	Ile	Gln 50	Arg	Arg	Gly	Arg	Asp 55	Ala	Trp	Gly	Pro	Gly 60	Asn	Ser	Ala	Arg
	Asp 65	Val	Leu	Arg		Arg 70.	Ala	Pro	Arg	Glu	Glu 75	Gln	Gly	Ala	Ala	Phe 80
	Leu	Ala	Gly	Pro	Ser 85	Trp	Asp	Leu	Pro	Ala 90	Ala	Pro	Gly	Arg	Asp 95	Pro
20	Ala	Ala	Gly	Arg 100	Gly	Ala	Glu	Ala	Ser 105	Ala	Ala	Gly	Pro	Pro 110	Gly	Pro
	Pro	Thr	Arg 115	Pro	Pro	Gly	Pro	Trp 120	Arg	Trp	Lys	Gly	Ala 125	Arg	Gly	Gln
25	Glu	Pro 130	Ser	Glu	Thr	Leu	Gly 135	Arg	Gly	Asn	Pro	Thr 140	Ala	Leu	Gln	Leu
	Phe 145	Leu	Gln	Ile	Ser	Glu 150	Glu	Glu	Glu	Lys	Gly 155	Pro	Arg	Gly	Ala	Gly 160
	Ile	Ser	Gly	Arg	Ser 165	Gln	Glu	Gln	Ser	Val 170	Lys	Thr	Val	Pro	Gly 175	Ala
30	Ser	Asp	Leu	Phe 180	Tyr	Trp	Pro	Arg	Arg 185	Ala	Gly	Lys	Leu	Gln 190	Gly	Ser
	His	His	Lys 195	Pro	Leu	Ser	Lys	Thr 200	Ala	Asn	Gly	Leu	Ala 205	Gly	His	Glu
35	Gly	Trp 210	Thr	Ile	Ala	Leu	Pro 215	Gly	Arg	Ala	Leu	Ala 220	Gln	Asn	Gly	Ser
	Leu 225	Gly	Glu	Gly	Ile	His 230	Glu	Pro	Gly	Gly	Pro 235	Arg	Arg	Gly	Asn	Ser 240

	Thr	Asn	Arg	Arg	Val 245	Arg	Leu	Lys	Asn	Pro 250	Phe	Tyr	Pro	Leu	Thr 255	Gln
•	Glu	Ser	Tyr	Gly 260	Ala	Tyr	Ala	Val	Met 265	Cys	Leu	Ser	Val	Val 270	Ile	Phe
5	Gly	Thr	Gly 275	Ile	Ile	Gly	Asn	Leu 280	Ala	Val	Met	Ser	Ile 285	Val	Cys	His
	Asn	Tyr 290	Tyr	Met	Arg	Ser	Ile 295	Ser	Asn	Ser	Leu	Leu 300	Ala	Asn	Leu	Ala
10	Phe 305	Trp	Asp	Phe	Leu	Ile 310	Ile	Phe	Phe	Cys	Leu 315	Pro	Leu	Val	Ile	Phe 320
	His	Glu	Leu	Thr	Lys 325	Lys	Trp	Leu	Leu	Glu 330	Asp	Phe	Ser	Cys	Lys 335	Ile
,	Val	Pro	Tyr	Ile 340	Glu	Val	Ala	Ser	Leu 345	Gly	Val	Thr	Thr	Phe 350	Thr	Leu
15	Cys	Ala	Leu 355	Cys	Ile	Asp	Arg	Phe 360	Arg	Ala	Ala	Thr	Asn 365	Val	Gln	Met
	Tyr	Tyr 370	Glu	Met	Ile	Glu	Asn 375	Cys	Ser	Ser	Thr	Thr 380	Ala	Lys	Leu	Ala
20	Val 385	Ile	Trp	Val	Gly	Ala 390	Leu	Leu	Leu	Ala	Leu 395	Pro	Glu	Val	Val	Leu 400
	Arg	Gln	Leu	Ser	Lys 405	Glu	Asp	Leu	Gly	Phe 410	Ser	Gly	Arg	Ala	Pro 415	Ala
	Glu	Arg	Cys	Ile 420	Ile	Lys	Ile	Ser	Pro 425	Asp	Leu	Pro	Asp	Thr 430	Ile	Tyr
25	Val	Leu	Ala 435	Leu	Thr	Tyr	Asp	Ser 440	Ala	Arg	Leu	Trp	Trp 445	Tyr	Phe	Gly
	Cys	Tyr 450	Phe	Cys	Leu	Pro	Thr 455	Leu	Phe	Thr	Ile	Thr 460	Cys	Ser	Leu	Val
30	Thr 465	Ala	Arg	Lys	Ile	Arg 470	Lys	Ala	Glu	Lys	Ala 475	Cys	Thr	Arg	Gly	Asn 480
	-	Arg			485					490					495	
		Thr		500					505					510		
35		Val	515					520					525			
	Leu	Leu	Asn	Ile	Ile	Ser	Gln	Phe	Leu	Leu	Phe	Phe	Lys	Ser	Cys	Val

			530		• • •			535		•	•.	•	540					
	٠	Thr 545	Pro	Val	Leu	Leu	Phe 550	Cys	Leu	Cys	Lys	Pro 555	Phe	Ser	Arg	Ala	Phe 560	
5	e.	Met	Glu	Cys	Cys	Cys 565	Cys	Cys	Cys	Glu	Glu 570	Cys	Ile	Ģln	Lys	Ser 575	Ser	
		Thr	Val	Thr	Ser 580	Asp	Asp	Asn	Asp	Asn 585	Glu	Tyr	Thr	Thr	Glu 590	Leu	Glu	
	٠.	Leu	Ser	Pro 595	Phe	Ser			Arg 600		Glu	Met	Ser	Thr 605	Phe	Ala	Ser	
10		Val	Gly 610	Thr	His	Cys												
	(94)	•	•		٠										٠.			
15		(1)	(A) (B) (C)	LEN TYP STR	CE CH NGTH: PE: 1 RANDE	34 nucle	base ic a	e pai cid singl	irs									·
			(D)	TOE	POLOG	Y: 1	inea	ır										·
		(ii	.) MC	LECU	TE 1	YPE:	DNA	(ge	enomi	.c)								
		(xi	.) SE	QUEN	ICE I	ESCR	IPTI	ON:	SEQ	ID N	10:93	:						· ·
20	CAGA	TTCA	G AG	AAAA	LAAAG	TGA	LATAI	GGT	TTTT	•								34
	(95)	INFO	RMAT	ION	FOR	SEQ	ID N	: IO:94	l :			•	: .					
-	•	(i)	(A)	LEN	E CH IGTH: PE: n	32	base	pai										
25		•	(C)	STR	ANDE	DNES	S: s	ingl	.e					٠				
		(ii) MO	LECU	LE T	YPE:	DNA	(ge	nomi	c)								
		(x	i) s	EQUE	NCE	DESC	RIPT	'ION :	SEQ	ID	NO : 9	4:						
	TTGGA	TCCC	T GG	TGCA	TAAC	TAA	TGAA	AGA	AT				• ,		·			32
30	(96)	INFO	RMAT	ON	FOR	SEQ	ID N	O:95	:									
35		(i)	(A) (B) (C)	LEN TYP STR	E CH GTH: E: n ANDE	124 ucle DNES	8 ba ic a S: s	se p cid ingl	airs	•								
		(ii) MO	LECU	LE T	YPE:	DNA	(ge	nomi	c)								

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

		ATGGTTTTTG	CTCACAGAAT	GGATAACAGC	AAGCCACATT	TGATTATTCC	TACACTTCTG	60
		GTGCCCCTCC	AAAACCGCAG	CTGCACTGAA	ACAGCCACAC	CTCTGCCAAG	CCAATACCTG	120
		ATGGAATTAA	GTGAGGAGCA	CAGTTGGATG	AGCAACCAAA	CAGACCTTCA	CTATGTGCTG	180
,	5	AAACCCGGGG	AAGTGGCCAC	AGCCAGCATC	TTCTTTGGGA	TTCTGTGGTT	GTTTTCTATC	240
		TTCGGCAATT	CCCTGGTTTG	TTTGGTCATC	CATAGGAGTA	GGAGGACTCA	GTCTACCACC	300
		AACTACTTTG	TGGTCTCCAT	GGCATGTGCT	GACCTTCTCA	TCAGCGTTGC	CAGCACGCCT	360
		TTCGTCCTGC	TCCAGTTCAC	CACTGGAAGG	TGGACGCTGG	GTAGTGCAAC	GTGCAAGGTT	420
		GTGCGATATT	TTCAATATCT	CACTCCAGGT	GTCCAGATCT	ACGTTCTCCT	CTCCATCTGC	480
	10	ATAGACCGGT	TCTACACCAT	CGTCTATCCT	CTGAGCTTCA	AGGTGTCCAG	AGAAAAAGCC	540
		AAGAAAATGA	TTGCGGCATC	GTGGATCTTT	GATGCAGGCT	TTGTGACCCC	TGTGCTCTTT	600
		TTCTATGGCT	CCAACTGGGA	CAGTCATTGT	AACTATTTCC	TCCCCTCCTC	TTGGGAAGGC	660
		ACTGCCTACA	CTGTCATCCA	CTTCTTGGTG	GGCTTTGTGA	TTCCATCTGT	CCTCATAATT	720
t.		TTATTTTACC	AAAAGGTCAT	AAAATATATT	TGGAGAATAG	GCACAGATGG	CCGAACGGTG	780
-	15	AGGAGGACAA	TGAACATTGT	CCCTCGGACA	AAAGTGAAAA	CTATCAAGAT	GTTCCTCATT	840
		TTAAATCTGT	TGTTTTTGCT	CTCCTGGCTG	CCTTTTCATG	TAGCTCAGCT	ATGGCACCCC	900
		CATGAACAAG	ACTATAAGAA	AAGTTCCCTT	GTTTTCACAG	CTATCACATG	GATATCCTTT	960
		AGTTCTTCAG	CCTCTAAACC	TACTCTGTAT	TCAATTTATA	ATGCCAATTT	TCGGAGAGGG	1020
		ATGAAAGAGA	CTTTTTGCAT	GTCCTCTATG	AAATGTTACC	GAAGCAATGC	CTATACTATC	1080
	20	ACAACAAGTT	CAAGGATGGC	СААААААА	TACGTTGGCA	TTTCAGAAAT	CCCTTCCATG	1140
		GCCAAAACTA	TTACCAAAGA	CTCGATCTAT	GACTCATTTG	ACAGAGAAGC	CAAGGAAAAA	1200
		AAGCTTGCTT	GGCCCATTAA	CTCAAATCCA	CCAAATACTT	TTGTCTAA		1248

(97) INFORMATION FOR SEQ ID NO:96:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 415 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- 30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

				-,	٠		•	•				•				
	Met 1	Val	Phe	Ala	His 5	Arg	Met	Asp	Asn	Ser 10	Lys	Pro	His	Leu	Ile 15	Ile
	Pro	Thr	Leu	Leu 20	Val	Pro	Leu	Gln	Asn 25	Arg	Ser	Cys	Thr	Glu 30	Thr	Ala
5	Thr	Pro	Leu 35	Pro	Ser	Gln	Tyr	Leu 40	Met	Glu	Leu	Ser	Glu 45	Glu	His	Ser
	Trp	Met 50	Ser	Asn	Gln	Thr	Asp 55	Leu	His	Tyr	Val	Leu 60	Lys	Pro	Gly	Glu
10	Val 65	Ala	Thr	Ala	Ser	Ile 70	Phe	Phe	Gly	Île	Leu 75	Trp	Leu	Phe	Ser	Ile 80
	Phe	Gly	Asn	Ser	Leu 85	Val	Cys	Leu	Val	Ile 90	His	Arg	Ser	Arg	Arg 95	Thr
	Gln	Ser	Thr	Thr 100	Asn	Tyr	Phe	Val	Val 105	Ser	Met	Ala	Cys	Ala 110	Asp	Leu
15	Leu	Ile	Ser 115	Val	Ala	Ser	Thr	Pro 120	Phe	Val	Leu	Leu	Gln 125	Phe	Thr	Thr
	Gly	Arg 130	Trp	Thr	Leu	Gly	Ser 135	Ala	Thr	Cys	Lys	Val 140	Val	Arg	Tyr	Phe
20	Gln 145	Tyr	Leu	Thr	Pro	Gly 150	Val	Gln	Ile	Tyr	Val 155	Leu	Leú	Ser	Ile	Cýs 160
	Ile	Asp	Arg	Phe	Tyr 165	Thr	Ile	Val	Tyr	Pro 170	Leu	Ser	Phe	Lys	Val 175	
	Arg	Glu	Lys	Ala 180	Lys	Lys	Met	Ile	Ala 185	Ala	Ser	Trp	Ile	Phe 190	Asp	Ala
25	Gly	Phe	Val 195	Thr	Pro	Val	Leu	Phe 200	Phe	туr	Gly	Ser	Asn 205	Trp	Asp	Ser
	His	Cys 210	Asn	Tyr	Phe	Leu	Pro 215	Ser	Ser	Trp	Glu	Gly 220	Thr	Ala	Tyr	Thr
30	Val 225	Ile	His	Phe	Leu	Val 230	Gly	Phe	Val	Ile	Pro 235	Ser	Val	Leu	Ile	Ile 240
	Leu	Phe	Tyr	Gln	Lys 245	Val	Ile	Lys	Tyr	Ile 250	Trp	Arg	Ile	Gly	Thr 255	Asp
	Gly	Arg	Thr	Val 260	Arg	Arg	Thr	Met	Asn 265	Ile	Val	Pro	Arg	Thr 270	Lys	Val
35	Lys	Thr	Ile 275	Lys	Met	Phe	Leu	Ile 280	Leu	Asn	Leu	Leu	Phe 285	Leu	Leu	Ser

		Trp	Leu 290	Pro	Phe	His	Val	Ala 295	Gln	Leu	Trp	His	Pro 300	His	Glu	Gln	Asp
		Tyr 305	Lys	Lys	Ser	Ser	Leu 310	Val	Phe	Thr	Ala	Ile 315	Thr	Trp	Ile	Ser	Phe 320
5		Ser	Ser	Ser	Ala	Ser 325	Lys	Pro	Thr	Leu	Tyr 330	Ser	Ile	Tyr	Asn	Ala 335	Asn
		Phe	Arg	Arg	Gly 340	Met	Lys	Glu	Thr	Phe 345	Cys	Met	Ser	Ser	Met 350	Lys	Cys
10		Tyr	Arg	Ser 355	Asn	Ala	Tyr	Thr	Ile 360	Thr	Thr	Ser		Arg .365	Met	Ala	Lys
		Lys	Asn 370	Tyr	Val	Gly	Ile	Ser 375	Glu	Ile	Pro	Ser	Met 380	Ala	Lys	Thr	Ile
	·	Thr 385	Lys	Asp	Ser	Ile	Tyr 390	Asp	Ser	Phe	Asp	Arg 395	Glu	Ala	Lys		Lys 400
15		Lys	Leu	Ala	Trp	Pro 405	Ile	Asn	Ser		Pro 410	Pro	Asn	Thr		Val 415	
	(98)	INFO	RMAT	NOI	FOR	SEQ	ID N	10:97	:							,	
20	(98) INFORMATION FOR SEQ ID NO:97: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear																
		(ii) MO	LECU	LE T	YPE:	DNA	(ge:	nomi	c)				•			
		(x:	i) S	EQUE	NCE I	DESC	RIPT	ION:	SEQ	ID 1	NO : 9'	7:			٠		
25	GGAAA	GCTT	A AC	GATC	CCCA	GGA	GCAA	CAT									30
	(99)	INFO	RMAT:	ION I	FOR :	SEQ :	ID N	D:98	:						•		
30		(i)	(A) (B) (C)	TYPI STRA	E CHI GTH: E: nu ANDEI OLOGY	31 l cle: ONES	oase ic ad 8: s:	pair cid ingle	rs								
		(ii)	MOI	ECUI	E T	PE:	DNA	(ger	omic	:)							•
		(xi)	SEC	OUENC	E DE	ESCRI	PTIC	N: S	SEQ I	D NO	:98:						
	CTGGG	ATCCT	' ACC	AGAG	CAT	TTTI	CACA	CA G	}								31
35	(100)	INFO	RMAT	ON	FOR	SEQ	ID N	IO: 99):								
		(i)	SEQU	ENCE	CHA	RACT	ERIS	TICS	:		,						

- (A) LENGTH: 1842 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- 5 (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

	ATGGGGCCCA	CCCTAGCGGT	TCCCACCCCC	TATGGCTGTA	TTGGCTGTAA	GCTACCCCAG	60
	CCAGAATACC	CACCGGCTCT	AATCATCTTT	ATGTTCTGCG	CGATGGTTAT	CACCATCGTT	120
	GTAGACCTAA	TCGGCAACTC	CATGGTCATT	TTGGCTGTGA	CGAAGAACAA	GAAGCTCCGG	180
10	AATTCTGGCA	ACATCTTCGT	GGTCAGTCTC	TCTGTGGCCG	ATATGCTGGT	GGCCATCTAC	240
	CCATACCCTT	TGATGCTGCA	TGCCATGTCC	ATTGGGGGCT	GGGATCTGAG	CCAGTTACAG	300
	TGCCAGATGG	TCGGGTTCAT	CACAGGGCTG	AGTGTGGTCG	GCTCCATCTT	CAACATCGTG	360
	GCAATCGCTA	TCAACCGTTA	CTGCTACATC	TGCCACAGCC	TCCAGTACGA	ACGGATCTTC	420
	AGTGTGCGCA	ATACCTGCAT	CTACCTGGTC	ATCACCTGGA	TCATGACCGT	CCTGGCTGTC	480
15	CTGCCCAACA	TGTACATTGG	CACCATCGAG	TACGATCCTC	GCACCTACAC	CTGCATCTTC	540
	AACTATCTGA	ACAACCCTGT	CTTCACTGTT	ACCATCGTCT	GCATCCACTT	CGTCCTCCCT	600
	CTCCTCATCG	TGGGTTTCTG	CTACGTGAGG	ATCTGGACCA	AAGTGCTGGC	GGCCCGTGAC	660
	CCTGCAGGGC	AGAATCCTGA	CAACCAACTT	GCTGAGGTTC	GCAATTTTCT	AACCATGTTT	720
	GTGATCTTCC	TCCTCTTTGC	AGTGTGCTGG	TGCCCTATCA	ACGTGCTCAC	TGTCTTGGTG	780
20	-GCTGTCAGTC	CGAAGGAGAT	GGCAGGCAAG	ATCCCCAACT	GGCTTTATCT	TGCAGCCTAC	840
	TTCATAGCCT	ACTTCAACAG	CTGCCTCAAC	GCTGTGATCT	ACGGGCTCCT	CAATGAGAAT	900
	TTCCGAAGAG	AATACTGGAC	CATCTTCCAT	GCTATGCGGC	ACCCTATCAT	ATTCTTCCCT	960
	GGCCTCATCA	GTGATATTCG	TGAGATGCAG	GAGGCCCGTA	CCCTGGCCCG	CGCCCGTGCC	1020
	CATGCTCGCG	ACCAAGCTCG	TGAACAAGAC	CGTGCCCATG	CCTGTCCTGC	TGTGGAGGAA	1080
25	ACCCCGATGA	ATGTCCGGAA	TGTTCCATTA	CCTGGTGATG	CTGCAGCTGG	CCACCCCGAC	1140
	CGTGCCTCTG	GCCACCCTAA	GCCCCATTCC	AGATCCTCCT	CTGCCTATCG	CAAATCTGCC	1200
	TCTACCCACC	ACAAGTCTGT	CTTTAGCCAC	TCCAAGGCTG	CCTCTGGTCA	CCTCAAGCCT	1260
	GTCTCTGGCC	ACTCCAAGCC	TGCCTCTGGT	CACCCCAAGT	CTGCCACTGT	CTACCCTAAG	1320
	CCTGCCTCTG	TCCATTTCAA	GGGTGACTCT	GTCCATTTCA	AGGGTGACTC	TGTCCATTTC	1380

AAGCCTGACT CTGTTCATTT CAAGCCTGCT TCCAGCAACC CCAAGCCCAT CACTGGCCAC 1440 CATGTCTCTG CTGGCAGCCA CTCCAAGTCT GCCTTCAGTG CTGCCACCAG CCACCCTAAA 1500 CCCATCAAGC CAGCTACCAG CCATGCTGAG CCCACCACTG CTGACTATCC CAAGCCTGCC 1560 ACTACCAGCC ACCCTAAGCC CGCTGCTGCT GACAACCCTG AGCTCTCTGC CTCCCATTGC 1620 CCCGAGATCC CTGCCATTGC CCACCCTGTG TCTGACGACA GTGACCTCCC TGAGTCGGCC 1680 TCTAGCCCTG CCGCTGGGCC CACCAAGCCT GCTGCCAGCC AGCTGGAGTC TGACACCATC 1740 GCTGACCTTC CTGACCCTAC TGTAGTCACT ACCAGTACCA ATGATTACCA TGATGTCGTG 1800 GTTGTTGATG TTGAAGATGA TCCTGATGAA ATGGCTGTGT GA 1842 (101) INFORMATION FOR SEQ ID NO:100: (i) SEQUENCE CHARACTERISTICS:. (A) LENGTH: 613 amino acids (B) TYPE: amino acid

- 10
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
- 15 (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

Met Gly Pro Thr Leu Ala Val Pro Thr Pro Tyr Gly Cys Ile Gly Cys 10

Lys Leu Pro Gln Pro Glu Tyr Pro Pro Ala Leu Ile Ile Phe Met Phe 20

> Cys Ala Met Val Ile Thr Ile Val Val Asp Leu Ile Gly Asn Ser Met 40

> Val Ile Leu Ala Val Thr Lys Asn Lys Lys Leu Arg Asn Ser Gly Asn

Ile Phe Val Val Ser Leu Ser Val Ala Asp Met Leu Val Ala Ile Tyr 25 65 75

Pro Tyr Pro Leu Met Leu His Ala Met Ser Ile Gly Gly Trp Asp Leu

Ser Gln Leu Gln Cys Gln Met Val Gly Phe Ile Thr Gly Leu Ser Val 30 105

> Val Gly Ser Ile Phe Asn Ile Val Ala Ile Ala Ile Asn Arg Tyr Cys 115 120

Tyr Ile Cys His Ser Leu Gln Tyr Glu Arg Ile Phe Ser Val Arg Asn 130 135

	Thr 145	Cys	Ile	Tyr	Leu	Val 150	Ile	Thr	Trp	Ile	Met 155	Thr	Val	Leu	Ala	Val 160
	Leu	Pro	Asn	Met	Tyr 165	Ile	Gly	Thr	Ile	Glu 170	Tyr	Asp	Pro	Arg	Thr 175	Tyr
5	Thr	Cys	Ile	Phe 180	Asn	Tyr	Leu	Asn	Asn 185	Pro	Val	Phe	Thr	Val 190	Thr	Ile
	Val	Cys	Ile 195	His	Phe	Val	Leu	Pro 200	Leu	Leu	Ile	Val	Gly 205	Phe	Cys	Tyr
10	Val	Arg 210	Ile	Trp	Thr	Lys	Val 215	Leu	Ala	Ala	Arg	Asp 220	Pro	Ala	Gly	Gln
	Asn 225	Pro	Asp	Asn	Gln	Leu 230	Ala	Glu	Val	Arg	Asn 235	Phe	Leu	Thr	Met	Phe 240
	Val	Ile	Phe	Leu	Leu 245	Phe	Ala	Val	Cys	Trp 250	Cys	Pro	Ile	Asn	Val 255	Leu
15	Thr	Val	Leu	Val 260	Ala	Val	Ser	Pro	Lys 265	Glu	Met	Ala	Gly	Lys 270	Ile	Pro
	Asn	Trp	Leu 275	Tyr	Leu	Ala	Ala	Tyr 280	Phe	Ile	Ala	Tyr	Phe 285	Asn	Ser	Cys
20	Leu	Asn 290	Ala	Val	Ile	Tyr	Gly 295	Leu	Leu	Asn	Glu	Asn 300	Phe	Arg	Arg	Glu
	Tyr 305	Trp	Thr	Ile	Phe	His 310	Ala	Met	Arg	His	Pro 315	Ile	Ile	Phe	Phe	Pro 320
	Gly	Leu	Ile	Ser	Asp 325	Ile	Arg	Glu	Met	Gln 330	Glu	Ala	Arg	Thr	Leu 335	Ala
25	Arg	Ala	Arg	Ala 340	His	Ala	Arg	Asp	Gln 345	Ala	Arg	Glu	Gln	Asp 350	Arg	Ala
	His	Ala	Cys 355	Pro	Ala	Val	Glu	Glu 360	Thr	Pro	Met	Asn	Val 365	Arg	Asn	Val
30	Pro	Leu 370	Pro	Gly	Asp	Ala	Ala 375	Ala	Gly	His	Pro	Asp 380	Arg	Ala	Ser	Gly
	His 385	Pro	Lys	Pro	His	Ser 390	Arg	Ser	Ser	Ser	Ala 395	Tyr	Arg	Lys	Ser	Ala 400
	Ser	Thr	His	His	Lys 405	Ser	Val	Phe	Ser	His 410	Ser	Lys	Ala	Ala	Ser 415	Gly
35	His	Leu	Lys	Pro 420	Val	Ser	Gly	His	Ser 425	Lys	Pro	Ala	Ser	Gly 430	His	Pro
	Lys	Ser	Ala	Thr	Val	Tyr	Pro	Lys	Pro	Ala	Ser	Val	His	Phe	Lys	Gly

				435					440	•				445				
		Asp	Ser 450	Val	His	Phe	Lys	Gly 455		Ser	Val	His	Phe 460		Pro	Asp	Ser	•
5		Val 465	His	Phe	Lys	Pro	Ala 470	Ser	Ser	Asn	Pro	Lys 475	Pro	Ile	Thr	Gly	His 480	
		His	Val	Ser	Ala	Gly 485	Ser	His	Ser	Lys	Ser 490	Ala	Phe	Ser	Ala	Ala 495	Thr	
		Ser	His	Pro	Lys 500	Pro	Ile	Lys	Pro	Ala 505	Thr	Ser	His	Ala	Glu 510	Pro	Thr	
10		Thr	Ala	Asp 515	Tyr	Pro	Lys	Pro	Ala 520	Thr	Thr	Ser	His	Pro 525	Lys	Pro	Ala	
		Ala	Ala 530	Asp	Asn	Pro	Glu	Leu 535	Ser	Ala	Ser	His	Cys 540	Pro	Glu	Ile	Pro	
15		Ala 545	Ile	Ala	His	Pro	Val 550	Ser	Asp	Asp	Ser	Asp 555	Leu	Pro	Glu	Ser	Ala 560	
		Ser	Ser	Pro	Ala	Ala 565	Gly	Pro	Thr	Lys	Pro 570	Ala	Ala	Ser	Gln	Leu 575	Glu	
		Ser	Asp	Thr	Ile 580	Ala	Asp	Leu	Pro	Asp 585	Pro	Thr	Val	Val	Thr 590	Thr	Ser	
20	•	Thr	Asn	Asp 595	Tyr	His	Asp	Val	Val 600	Val	Val	Asp	Val	Glu 605	Asp	Asp	Pro	
		Asp	Glu 610	Met	Ala	Val												
	(102)	INF	ORMA	TION	FOF	SEQ	ID	NO:1	01:									
25		(i)	(A) (B) (C)	LEN TYP STR	GTH: E: n ANDE	IARAC 32 Iucle DNES Y: 1	base ic a S: s	pai cid ingl	rs									
30		(ii) MO	LECU	LE I	YPE:	DNA	. (ge	nomi	c).								
						ESCR				·		1:						
	TCCAA																	32
	(103)	INF	orma	TION	FOR	SEQ	ID	NO:1	02:									
35		(i)	(A) (B)	LEN TYP	GTH: E: n	ARAC 30 1 ucle	base ic a	pai cid	rs									

30

81

/DI	TOPOLOGY:	linear

- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

CGTGAATTCC AAGAATTTAC AATCCTTGCT

30

(104) INFORMATION FOR SEQ ID NO:103:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1548 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
- 10 (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

ATGGGACATA ACGGGAGCTG GATCTCTCCA AATGCCAGCG AGCCGCACAA CGCGTCCGGC 60 GCCGAGGCTG CGGGTGTGAA CCGCAGCGCG CTCGGGGAGT TCGGCGAGGC GCAGCTGTAC 120 15 CGCCAGTTCA CCACCACCGT GCAGGTCGTC ATCTTCATAG GCTCGCTGCT CGGAAACTTC 180 ATGGTGTTAT GGTCAACTTG CCGCACAACC GTGTTCAAAT CTGTCACCAA CAGGTTCATT 240 AAAAACCTGG CCTGCTCGGG GATTTGTGCC AGCCTGGTCT GTGTGCCCTT CGACATCATC 300 CTCAGCACCA GTCCTCACTG TTGCTGGTGG ATCTACACCA TGCTCTTCTG CAAGGTCGTC 360 AAATTTTTGC ACAAAGTATT CTGCTCTGTG ACCATCCTCA GCTTCCCTGC TATTGCTTTG 420 20 GACAGGTACT ACTCAGTCCT CTATCCACTG GAGAGGAAAA TATCTGATGC CAAGTCCCGT 480 GAACTGGTGA TGTACATCTG GGCCCATGCA GTGGTGGCCA GTGTCCCTGT GTTTGCAGTA 540 ACCAATGTGG CTGACATCTA TGCCACGTCC ACCTGCACGG AAGTCTGGAG CAACTCCTTG 600 GGCCACCTGG TGTACGTTCT GGTGTATAAC ATCACCACGG TCATTGTGCC TGTGGTGGTG 660 GTGTTCCTCT TCTTGATACT GATCCGACGG GCCCTGAGTG CCAGCCAGAA GAAGAAGGTC 720 ATCATAGCAG CGCTCCGGAC CCCACAGAAC ACCATCTCTA TTCCCTATGC CTCCCAGCGG 780 GAGGCCGAGC TGCACGCCAC CCTGCTCTCC ATGGTGATGG TCTTCATCTT GTGTAGCGTG 840 CCCTATGCCA CCCTGGTCGT CTACCAGACT GTGCTCAATG TCCCTGACAC TTCCGTCTTC 900 TTGCTGCTCA CTGCTGTTTG GCTGCCCAAA GTCTCCCTGC TGGCAAACCC TGTTCTCTTT 960 CTTACTGTGA ACAAATCTGT CCGCAAGTGC TTGATAGGGA CCCTGGTGCA ACTACACCAC 1020 CGGTACAGTC GCCGTAATGT GGTCAGTACA GGGAGTGGCA TGGCTGAGGC CAGCCTGGAA

CCCAGCATAC GCTCGGGTAG CCAGCTCCTG GAGATGTTCC ACATTGGGCA GCAGCAGATC 1140

TTTAAGCCCA CAGAGGATGA GGAAGAGAGT GAGGCCAAGT ACATTGGCTC AGCTGACTTC 1200

CAGGCCAAGG AGATATTTAG CACCTGCCTC GAGGGAGAGC AGGGGCCACA GTTTGCGCCC 1260

TCTGCCCCAC CCCTGAGCAC AGTGGACTCT GTATCCCAGG TGGCACCGGC AGCCCCTGTG 1320

GAACCTGAAA CATTCCCTGA TAAGTATTCC CTGCAGTTTG GCTTTGGGCC TTTTGAGTTG 1380

CCTCCTCAGT GGCTCTCAGA GACCCGAAAC AGCAAGAAGC GGCTGCTTCC CCCCTTGGGC 1440

AACACCCCAG AAGAGCTGAT CCAGACAAAG GTGCCCAAGG TAGGCAGGGT GGAGCGGAAG 1500

ATGAGCAGAA ACAATAAAGT GAGCATTTTT CCAAAGGTGG ATTCCTAG 1548

(105) INFORMATION FOR SEQ ID NO:104:

10 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 515 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- 15 (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:

Met Gly His Asn Gly Ser Trp Ile Ser Pro Asn Ala Ser Glu Pro His 1 5 10 15

Asn Ala Ser Gly Ala Glu Ala Ala Gly Val Asn Arg Ser Ala Leu Gly 20 25 30

Glu Phe Gly Glu Ala Gln Leu Tyr Arg Gln Phe Thr Thr Thr Val Gln 35 40 45

Val Val Ile Phe Ile Gly Ser Leu Leu Gly Asn Phe Met Val Leu Trp 50 55 60

Ser Thr Cys Arg Thr Thr Val Phe Lys Ser Val Thr Asn Arg Phe Ile
65 70 75 80

Lys Asn Leu Ala Cys Ser Gly Ile Cys Ala Ser Leu Val Cys Val Pro 85 90 95

Phe Asp Ile Ile Leu Ser Thr Ser Pro His Cys Cys Trp Trp Ile Tyr 100 105 110

Thr Met Leu Phe Cys Lys Val Val Lys Phe Leu His Lys Val Phe Cys 115 120 125

Ser Val Thr Ile Leu Ser Phe Pro Ala Ile Ala Leu Asp Arg Tyr Tyr 130 140

	Ser 145	Val	Leu	Tyr	Pro	Leu 150	Glu	Arg	Lys	Ile	Ser 155		Ala	Lys	Ser	Arg 160
	Glu	Leu	Val	Met	Tyr 165	Ile	Trp	Ala	His	Ala 170		Val	Ala	Ser	Val 175	Pro
5	Val	Phe	Ala	Val 180	Thr	Asn	Val	Ala	Asp 185	Ile	Туr	Ala	Thr	Ser 190	Thr	Cys
• .	Thr	Glu	Val 195	Trp	Ser	Asn	Ser	Leu 200	Gly	His	Leu	Val	Tyr 205		Leu	Val
10	Tyr	Asn 210	Ile	Thr	Thr	Val	Ile 215	Val	Pro	Val	Val	Val 220	Val	Phe	Leu	Phe
	Leu 225	Ile	Leu	Ile	Arg	Arg 230	Ala	Leu	Ser	Ala	Ser 235	Gln	Lys	Lys	Lys	Val 240
	Ile	Ile	Ala	Ala	Leu 245	Arg	Thr	Pro	Gln	Asn 250	Thr	Ile	Ser	Ile	Pro 255	Tyr
15	Ala	Ser	Gln	Arg 260	Glu	Ala	Glu	Leu	His 265	Ala	Thr	Leu	Leu	Ser 270	Met	Val
	Met	Val	Phe 275	Ile	Leu	Cys	Ser	Val 280	Pro	Tyr	Ala	Thr	Leu 285	Val	Val	Tyr
20	Gln	Thr 290	Val	Leu	Asn	Val	Pro 295	Asp	Thr	Ser	Val	Phe 300	Leu	Leu	Leu	Thr
	Ala 305	Val	Trp	Leu	Pro	Lys 310	Val	Ser	Leu	Leu	Ala 315	Asn	Pro	Val	Leu	Phe 320
	Leu	Thr	Val	Asn	Lys 325	Ser	Val	Arg	Lys	Cys 330	Leu	Ile	Gly	Thr	Leu 335	Val
25	Gln	Leu	His	His 340	Arg	Tyr	Ser	Arg	Arg 345	Asn	Val	Val	Ser	Thr 350	Gly	Ser
	Gly	Met	Ala 355	Glu	Ala	Ser	Leu	Glu 360	Pro	Ser	Ile	Arg	Ser 365	Gly	Ser	Gln
30	Leu	Leu 370	Glu	Met	Phe	His	Ile 375	Gly	Gln	Gln	Gln	Ile 380	Phe	Lys	Pro	Thr
	Glu 385	Asp	Glu	Glu	Glu	Ser 390	Glu	Ala	Lys	Tyr	Ile 395	Gly	Ser	Ala	Asp .	Phe 400
	Gln	Ala	Lys	Glu	11e 405	Phe	Ser	Thr	Cys	Leu 410 _.	Glu	Gly	Glu	Gln	Gly 415	Pro
35	Gln	Phe	Ala	Pro 420	Ser	Ala	Pro	Pro	Leu 425	Ser	Thr	Val	Asp	Ser 430	Val	Ser
	Gln	Val	Ala	Pro	Ala	Ala	Pro	Val	Glu	Pro	Glu	Thr	Phe	Pro	Asp	Lys

				435	•				440					445			
		Tyr	Ser 450	Leu	Gln	Phe	Gly	Phe 455	Gly	Pro	Phe	Glu	Leu 460	Pro	Pro	Gln	Trp
5		Leu 465	Ser	Glu	Thr	Arg	Asn 470	Ser	Lys	Lys	Arg	Leu 475		Pro	Pro	Leu	Gly 480
	;	Asn	Thr	Pro	Glu	Glu 485	Leu	Ile	Gln	Thr	Lys 490	Val	Pro	Lys	Val	Gly 495	Arg
	,	Val	Glu	Arg	Lys 500	Met	Ser	Arg	Asn	Asn 505	Lys	Val	Ser	Ile	Phe 510	Pro	Lys
10	•	Val	Asp	Ser 515													
	(106)	INF	ORMA	TION	FOF	SEC) ID	NO:	L05 :						:	•	
15		(i)	(A)	LEN	IGTH :	IARAC 29	base	pai	rs					٠			
13			(C)	STR	ANDE	ucle DNES Y: 1	S: s	ingl		,						٠	
		(ii) MO	LECU	LE T	YPE:	DNA	. (ge	nomi	.c)				,			
		(xi) SE	QUEN	CE D	ESCR	IPTI	ON:	SEQ	ID N	0:10	5:			•		
20	GGAGAA								_		7,						29
	(107)								06 •								29
			SEQ (A)	UENC LEN	E CH GTH:	ARAC	TERI base	STIC pai	s:						• .		
25			(C)	STR	ANDE	ucle DNES	S: s	ingl	e								
			(D)	TOP	OLOG	Y: 1	inea	r									
		(ii)) MO	LECU	LE T	YPE:	DNA	(ge	nomi	c)							
		(xi)) SE	QUEN	CE D	ESCR	IPTI	ON:	SEQ	ID N	0:10	6:					
	GGAGGA	TCC	A GG	AAAC	CTTA	GGC	CGAG'	TCC									30
30	(108)	INFO	ORMA'	rion	FOR	SEQ	ID 1	NO:1	07:.								
35			(A) (B) (C) (D)	TYPE STREETOP	GTH: E: nv ANDEI OLOGY	ARAC: 1164 10le: ONESS 7: 1:	basic ac S: s:	se pa cid ingla	airs e								
		(11)	MOI	15.CUI	ים אר	rrE:	DNA	(gei	nomi	2)							

ANSUCCIO SWO COSSISSAT IN-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

	ATGAATCGGC	ACCATCTGCA	GGATCACTTT	CTGGAAATAG	ACAAGAAGAA	CTGCTGTGTG	60
	TTCCGAGATG	ACTTCATTGC	CAAGGTGTTG	CCGCCGGTGT	TGGGGCTGGA	GTTTATCTTT	120
	GGGCTTCTGG	GCAATGGCCT	TGCCCTGTGG	ATTTTCTGTT	TCCACCTCAA	GTCCTGGAAA	180
5	TCCAGCCGGA	TTTTCCTGTT	CAACCTGGCA	GTAGCTGACT	TTCTACTGAT	CATCTGCCTG	240
	CCGTTCGTGA	TGGACTACTA	TGTGCGGCGT	TCAGACTGGA	ACTTTGGGGA	CATCCCTTGC	300
	CGGCTGGTGC	TCTTCATGTT	TGCCATGAAC	CGCCAGGGCA	GCATCATCTT	CCTCACGGTG	360
	GTGGCGGTAG	ACAGGTATTT	CCGGGTGGTC	CATCCCCACC	ACGCCCTGAA	CAAGATCTCC	420
	AATTGGACAG	CAGCCATCAT	CTCTTGCCTT	CTGTGGGGCA	TCACTGTTGG	CCTAACAGTC	480
10	CACCTCCTGA	AGAAGAAGTT	GCTGATCCAG	AATGGCCCTG	CAAATGTGTG	CATCAGCTTC	540
	AGCATCTGCC	ATACCTTCCG	GTGGCACGAA	GCTATGTTCC	TCCTGGAGTT	CCTCCTGCCC	600
	CTGGGCATCA	TCCTGTTCTG	CTCAGCCAGA	ATTATCTGGA	GCCTGCGGCA	GAGACAAATG	660
	GACCGGCATG	CCAAGATCAA	GAGAGCCATC	ACCTTCATCA	TGGTGGTGGC	CATCGTCTTT	720
	GTCATCTGCT	TCCTTCCCAG	CGTGGTTGTG	CGGATCCGCA	TCTTCTGGCT	CCTGCACACT	780
15	TCGGGCACGC	AGAATTGTGA	AGTGTACCGC	TCGGTGGACC	TGGCGTTCTT	TATCACTCTC	840
	AGCTTCACCT	ACATGAACAG	CATGCTGGAC	CCCGTGGTGT	ACTACTTCTC	CAGCCCATCC	900
	TTTCCCAACT	TCTTCTCCAC	TTTGATCAAC	CGCTGCCTCC	AGAGGAAGAT	GACAGGTGAG	960
	CCAGATAATA	ACCGCAGCAC	GAGCGTCGAG	CTCACAGGGG	ACCCCAACAA	AACCAGAGGC	1020
	GCTCCAGAGG	CGTTAATGGC	CAACTCCGGT	GAGCCATGGA	GCCCTCTTA	TCTGGGCCCA	1080
20	ACCTCAAATA	ACCATTCCAA	GAAGGGACAT	TGTCACCAAG	AACCAGCATC	TCTGGAGAAA	1140
	CAGTTGGGCT	GTTGCATCGA	GTAA				1164

(109) INFORMATION FOR SEQ ID NO:108:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 387 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:

- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:
- 30 Met Asn Arg His His Leu Gln Asp His Phe Leu Glu Ile Asp Lys Lys

	1				_ 5	<i>,</i> · .				. 10), , ,			. •	15	i
	As	n Cy	/s Cy	s Va 20	l Ph	e Ar	g As	p As	p Ph 25	e Il	e Al	a Ly	's Va	l Le 30		o Pro
, 5	Va	l Le	u Gl 35	y Le	u Glı	u Ph	e Il	e Ph 40	e Gl	y Le	u Le	u Gl	y As 45		y Le	u Ala
	Le	u Tr 50	p Il	e Ph	e Cys	s Ph	e Hi: 55	s Le	u Ly	s Se	r Tr	р Ly 60	s Se	r Se	r Ar	g Ile
	Pho 65	e Le	u Ph	e As	n Lev	1 Ala 70	a Va.	l Ala	a As	p Ph	e Le [.] 75	u Le	u Il	e Il	е Су	s Leu 80
10	Pro	o Ph	e Va	l Me	t Asp 85	Туз	г Туг	r Val	l Ar	g Ar	g Se:	r Asj	p Tr	p Ası	n Phe 95	e Gly
	Ası	o Il	e Pro	100	s Arg	Let	ı Val	Lei	1 Phe		t Phe			Ası 110		g Gln
15	Gl	/ Se	r Ile 115	e Ile	e Phe	Leu	ı Thr	Val	. Va]	l Ala	a Val	. Asp	Arg 125		Phe	e Arg
	Val	. Va:	l His	Pro	His	His	Ala 135	Leu	ı Asr	Lys	: Ile	Ser 140		Trp	Thr	Ala
	Ala 145	ı Ile	e Ile	e Ser	Cys	Leu 150	Leu	Trp	Gly	' Ile	Thr 155		Gly	Leu	Thr	Val 160
20	His	Leu	ı Leu	Lys	Lys 165	Lys	Leu	Leu	Ile	Gln 170		Gly	Pro	Ala	Asn 175	
	Cys	Ile	e Ser	Phe 180	Ser	Ile	Cys	His	Thr 185	Phe	Arg	Trp	His	Glu 190	Ala	Met
25	Phe	Leu	Leu 195	Glu	Phe	Leu	Leu	Pro 200		Gly	Ile	Ile	Leu 205	Phe	Cys	Ser
	Ala	Arg 210	Ile	Ile	Trp	Ser	Leu 215	Arg	Gln	Arg	Gln	Met 220	Asp	Arg	His	Ala
	Lys 225	Ile	Lys	Arg	Ala	Ile 230	Thr	Phe	Ile	Met	Val 235	Val	Ala	Ile	Val	Phe 240
30	Val	Ile	Cys	Phe	Leu 245	Pro	Ser	Val	Val	Val 250	Arg	Ile	Arg	Ile	Phe 255	Trp
	Leu	Leu	His	Thr 260	Ser	Gly	Thr	Gln	Asn 265	Cys	Glu	Val	Tyr	Arg 270	Ser	Val
35	Asp	Leu	Ala 275	Phe	Phe	Ile	Thr	Leu 280	Ser	Phe	Thr	Tyr	Met 285	Asn	Ser	Met
	Leu	Asp 290	Pro	Val	Val '	Tyr	Tyr 295	Phe	Ser	Ser		Ser 300	Phe	Pro	Asn	Phe

BNSDOCID <WO 002212941 14.

	Phe Ser Thr Leu Ile Asn Arg Cys Leu Gln Arg Lys Met Thr Gly Glu 305 310 315 320	
	Pro Asp Asn Asn Arg Ser Thr Ser Val Glu Leu Thr Gly Asp Pro Asn 325 330 335	
5	Lys Thr Arg Gly Ala Pro Glu Ala Leu Met Ala Asn Ser Gly Glu Pro 340 345 350	
	Trp Ser Pro Ser Tyr Leu Gly Pro Thr Ser Asn Asn His Ser Lys Lys 355 360 365	
10	Gly His Cys His Gln Glu Pro Ala Ser Leu Glu Lys Gln Leu Gly Cys 370 375 380	
	Cys Ile Glu 385	
	(110) INFORMATION FOR SEQ ID NO:109:	
15	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 37 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
20	(iv) ANTI-SENSE: NO	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:	
	ACCATGGCTT GCAATGGCAG TGCGGCCAGG GGGCACT	37
	(111) INFORMATION FOR SEQ ID NO:110:	
25	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
30	(iv) ANTI-SENSE: YES	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:	
	CGACCAGGAC AAACAGCATC TTGGTCACTT GTCTCCGGC	3 9
	(112) INFORMATION FOR SEQ ID NO:111:	
35	(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 39 base pairs(B) TYPE: nucleic acid	

	(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
5	(iv) ANTI-SENSE: NO	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:	
	GACCAAGATG CTGTTTGTCC TGGTCGTGGT GTTTGGCAT	39
	(113) INFORMATION FOR SEQ ID NO:112:	
10	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 35 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
15	(iv) ANTI-SENSE: YES	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:	
	CGGAATTCAG GATGGATCGG TCTCTTGCTG CGCCT	35
	(114) INFORMATION FOR SEQ ID NO:113:	
20	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1212 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
25	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:	
	ATGGCTTGCA ATGGCAGTGC GGCCAGGGGG CACTTTGACC CTGAGGACTT GAACCTGACT	60
	GACGAGGCAC TGAGACTCAA GTACCTGGGG CCCCAGCAGA CAGAGCTGTT CATGCCCATC	120
	TGTGCCACAT ACCTGCTGAT CTTCGTGGTG GGCGCTGTGG GCAATGGGCT GACCTGTCTG	180
	GTCATCCTGC GCCACAAGGC CATGCGCACG CCTACCAACT ACTACCTCTT CAGCCTGGCC	240
30	GTGTCGGACC TGCTGGTGCT GCTGGTGGGC CTGCCCCTGG AGCTCTATGA GATGTGGCAC	300
	AACTACCCCT TCCTGCTGGG CGTTGGTGGC TGCTATTTCC GCACGCTACT GTTTGAGATG	360
	GTCTGCCTGG CCTCAGTGCT CAACGTCACT GCCCTGAGCG TGGAACGCTA TGTGGCCGTG	420
	GTGCACCCAC TCCAGGCCAG GTCCATGGTG ACGCGGGCCC ATGTGCGCCG AGTGCTTGGG	480

WO 00/22129 PCT/US99/23938

89

				07			
	GCCGTCTGGG	GTCTTGCCAT	GCTCTGCTCC	CTGCCCAACA	CCAGCCTGCA	CGGCATCCGG	54
	CAGCTGCACG	TGCCCTGCCG	GGGCCCAGTG	CCAGACTCAG	CTGTTTGCAT	GCTGGTCCGC	60
	CCACGGGCCC	TCTACAACAT	GGTAGTGCAG	ACCACCGCGC	TGCTCTTCTT	CTGCCTGCCC	66
	ATGGCCATCA	TGAGCGTGCT	CTACCTGCTC	ATTGGGCTGC	GACTGCGGCG	GGAGAGGCTG	720
5	CTGCTCATGC	AGGAGGCCAA	GGGCAGGGGC	TCTGCAGCAG	CCAGGTCCAG	ATACACCTGC	780
	AGGCTCCAGC	AGCACGATCG	GGGCCGGAGA	CAAGTGACCA	AGATGCTGTT	TGTCCTGGTC	840
	GTGGTGTTTG	GCATCTGCTG	GGCCCCGTTC	CACGCCGACC	GCGTCATGTG	GAGCGTCGTG	900
	TCACAGTGGA	CAGATGGCCT	GCACCTGGCC	TTCCAGCACG	TGCACGTCAT	CTCCGGCATC	960
	TTCTTCTACC	TGGGCTCGGC	GGCCAACCCC	GTGCTCTATA	GCCTCATGTC	CAGCCGCTTC	1020
10	CGAGAGACCT	TCCAGGAGGC	CCTGTGCCTC	GGGCCTGCT	GCCATCGCCT	CAGACCCCGC	1080
	CACAGCTCCC	ACAGCCTCAG	CAGGATGACC	ACAGGCAGCA	CCCTGTGTGA	TGTGGGCTCC	1140
	CTGGGCAGCT	GGGTCCACCC	CCTGGCTGGG	AACGATGGCC	CAGAGGCGCA	GCAAGAGACC	1200
	GATCCATCCT	-GA				,	1212
	(115) INFOR	MATION FOR	SEQ ID NO:1	14:			
15	(SEQUENCE CHA (A) LENGTH: (B) TYPE: and (C) STRANDEL (D) TOPOLOGY	403 amino a nino acid DNESS:	acids			
20	(ii)	MOLECULE TY	PE: protein	1			

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

Met Ala Cys Asn Gly Ser Ala Ala Arg Gly His Phe Asp Pro Glu Asp 1 5 10 15

Leu Asn Leu Thr Asp Glu Ala Leu Arg Leu Lys Tyr Leu Gly Pro Gln 25 25 30

Gln Thr Glu Leu Phe Met Pro Ile Cys Ala Thr Tyr Leu Leu Ile Phe 35 40 45

Val Val Gly Ala Val Gly Asn Gly Leu Thr Cys Leu Val Ile Leu Arg 50 55 60

His Lys Ala Met Arg Thr Pro Thr Asn Tyr Tyr Leu Phe Ser Leu Ala 65 70 75 80

Val Ser Asp Leu Leu Val Leu Leu Val Gly Leu Pro Leu Glu Leu Tyr 85 90 95

	Glu	Met	Trp	His 100	Asn	Tyr	Pro	Phe	Leu 105	Leu	Gly	Val	Gly	Gly 110	Cys	Tyr
	Phe	Arg	Thr 115	Leu	Leu	Phe	Glu	Met 120	Val	Cys	Leu	Ala	Ser 125		Leu	Asn
5	Val	Thr 130	Ala	Leu	Ser	Val	Glu 135	Arg	Tyr	Val	Ala	Val 140	Val	His	Pro	Leu
	Gln 145	Ala	Arg	Ser	Met	Val 150	Thr	Arg	Ala	His	Val 155	Arg	Arg	Val	Leu	Gly 160
10	Ala	Val	Trp	Gly	Leu 165	Ala	Met	Leu	Cys	Ser 170	Leu	Pro	Asn	Thr	Ser 175	Leu
	His	Gly	Ile	Arg 180	Gln	Leu	His	Val	Pro 185	Cys	Arg	Gly		Val 190		Asp
	Ser	Ala	Val 195	Cys	Met	Leu	Val	Arg 200	Pro	Arg	Ala	Leu	Tyr 205	Asn	Met	Val
15		210					215					220				Met
	225	Val				230					235					240
20		Leu			245					250					255	
		Tyr		260					265					270		
26		Lys	275					280		•			285	_		
25		Phe 290					295					300				
	305	Gly				310					315					320
30		Phe			325					330					335	
		Ser		340					345					350		
35		Cys	355					360					365			
<i>,</i>	Met	370					375					380				_
	vdl	His	P.LO	ביבע	u Tq	GT À	ASII	ASP	дīЛ	Pro	GIU	Ата	GIN	GIn	GLu	Thr

PCT/US99/23938

	385 390 395	400
	Asp Pro Ser	
	(116) INFORMATION FOR SEQ ID NO:115:	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
10	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:	
	GGAAGCTTCA GGCCCAAAGA TGGGGAACAT	30
	(117) INFORMATION FOR SEQ ID NO:116:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	y.
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:	•
	GTGGATCCAC CCGCGGAGGA CCCAGGCTAG	30
	(118) INFORMATION FOR SEQ ID NO:117:	
25	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1098 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:	
30	ATGGGGAACA TCACTGCAGA CAACTCCTCG ATGAGCTGTA CCATCGACCA TACCATCCAC	60
	CAGACGCTGG CCCCGGTGGT CTATGTTACC GTGCTGGTGG TGGGCTTCCC GGCCAACTGC	120
	CTGTCCCTCT ACTTCGGCTA CCTGCAGATC AAGGCCCGGA ACGAGCTGGG CGTGTACCTC	180
	TGCAACCTGA CGGTGGCCGA CCTCTTCTAC ATCTGCTCGC TGCCCTTCTG GCTGCAGTAC	240
	GTGCTGCAGC ACGACAACTG GTCTCACGGC GACCTGTCCT GCCAGGTGTG CGGCATCCTC	300
35	CTGTACGAGA ACATCTACAT CAGCGTGGGC TTCCTCTGCT GCATCTCCGT GGACCGCTAC	360

			•				
	CTGGCTGTGG	CCCATCCCTT	CCGCTTCCAC	CAGTTCCGGA	. CCCTGAAGGC	GGCCGTCGGC	420
	GTCAGCGTGG	TCATCTGGGC	CAAGGAGCTG	CTGACCAGCA	TCTACTTCCT	GATGCACGAG	480
	GAGGTCATCG	AGGACGAGAA	CCAGCACCGC	GTGTGCTTTG	AGCACTACCC	CATCCAGGCA	540
	TGGCAGCGCG	CCATCAACTA	CTACCGCTTC	CTGGTGGGCT	TCCTCTTCCC	CATCTGCCTG	600
5	CTGCTGGCGT	CCTACCAGGG	CATCCTGCGC	GCCGTGCGCC	GGAGCCACGG	CACCCAGAAG	660
	AGCCGCAAGG	ACCAGATCCA	GCGGCTGGTG	CTCAGCACCG	TGGTCATCTT	CCTGGCCTGC	720
	TTCCTGCCCT	ACCACGTGTT	GCTGCTGGTG	CGCAGCGTCT	GGGAGGCCAG	CTGCGACTTC	780
	GCCAAGGGCG	TTTTCAACGC	CTACCACTTC	TCCCTCCTGC	TCACCAGCTT	CAACTGCGTC	840
	GCCGACCCCG	TGCTCTACTG	CTTCGTCAGC	GAGACCACCC	ACCGGGACCT	GGCCCGCCTC	900
10	CGCGGGGCCT	GCCTGGCCTT	CCTCACCTGC	TCCAGGACCG	GCCGGGCCAG.	GGAGGCCTAC	960
	CCGCTGGGTG	CCCCGAGGC	CTCCGGGAAA	AGCGGGGCCC	AGGGTGAGGA	GCCCGAGCTG	1020
	TTGACCAAGC	TCCACCCGGC	CTTCCAGACC	CCTAACTCGC	CAGGGTCGGG	CGGGTTCCCC	1080
	ACGGGCAGGT	TGGCCTAG				•	1098
	(119) INFOR	RMATION FOR	SEQ ID NO:	118:	•		
15	, ,	SEQUENCE CHA (A) LENGTH: (B) TYPE: an	365 amino a	acids			

- 15

 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
- 20 (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

Met Gly Asn Ile Thr Ala Asp Asn Ser Ser Met Ser Cys Thr Ile Asp 5

His Thr Ile His Gln Thr Leu Ala Pro Val Val Tyr Val Thr Val Leu 25 25 30

> Val Val Gly Phe Pro Ala Asn Cys Leu Ser Leu Tyr Phe Gly Tyr Leu 40

> Gln Ile Lys Ala Arg Asn Glu Leu Gly Val Tyr Leu Cys Asn Leu Thr 50 55 60

30 Val Ala Asp Leu Phe Tyr Ile Cys Ser Leu Pro Phe Trp Leu Gln Tyr 65 70

> Val Leu Gln His Asp Asn Trp Ser His Gly Asp Leu Ser Cys Gln Val 90

		Cys	Gly	Ile	Leu 100	Leu	Tyr	Glu	Asn	Ile 105	Tyr	Ile	Ser	Val	Gly 110	Phe	Let
		Cys	Cys	Ile 115	Ser	Val	Asp	Arg	Tyr 120	Leu	Ala	Val	Ala	His 125	Pro	Phe	Arg
5		Phe	His 130	.Gln	Phe	Arg	Thr	Leu 135	Lys	Ala	Ala	Val	Gly 140	Val	Ser	Val	Val
		Ile 145	Trp	Ala	Lys	Glu	Leu 150	Leu	Thr	Ser	Ile	Tyr 155	Phe	Leu	Met	His	Glu 160
10		Glu	Val	Ile	Glu	Asp 165	Glu	Asn	Gln	His	Arg 170	Val	Cys	Phe	Glu	His 175	Tyr
		Pro	Ile	Gln	Ala 180	Trp	Gln	Arg	Ala	Ile 185	Asn	Tyr	Tyr	Arg	Phe 190	Leu	Val
		Gly	Phe	Leu 195	Phe	Pro	Ile	Cys	Leu 200	Leu	Leu	Ala	Ser	Tyr 205	Gln	Gly	Ile
15		Leu	Arg 210	Ala	Val	Arg	Arg	Ser 215	His	Gly	Thr	Gln	Lys 220	Ser	Arg	Lys	Asp
		Gln 225	Ile	Gln	Arg	Leu	Val 230	Leu	Ser	Thr	Val	Val 235	Ile	Phe	Leu	Ala	Cys 240
20		Phe	Leu	Pro	Tyr	His 245	Val	Leu	Leu	Leu	Val 250	Arg	Ser	Val	Trp	Glu 255	Ala
		Ser	Cys	Asp	Phe 260	Ala	Lys	Gly	Val	Phe 265	Asn	Ala	Tyr	His	Phe 270	Ser	Leu
		Leu	Leu	Thr 275	Ser	Phe	Asn	Cys	Val 280	Ala	Asp	Pro	Val	Leu 285	Tyr	Cys	Phe
25			Ser 290	Glu	Thr	Thr	His	Arg 295	Asp	Leu	Ala	Arg	Leu 300	Arg	Gly	Ala	Cys
		Leu 305	Ala	Phe	Leu	Thr	Cys 310	Ser	Arg	Thr	Gly	Arg 315	Ala	Arg	Glu	Ala	Tyr 320
30		Pro	Leu	Gly	Ala	Pro 325	Glu	Ala	Ser	Gly	Lys 330	Ser	Gly	Ala	Gln	Gly 335	Glu
	ı	Glu	Pro	Glu	Leu 340	Leu	Thr	Lys	Leu	His 345	Pro	Ala	Phe	Gln	Thr 350	Pro	Asn
		Ser	Pro	Gly 355	Ser	Gly	Gly	Phe	Pro 360	Thr	Gly	Arg	Leu	Ala 365			
35	(120)	INF	ORMA	TION	FOR	SEO	ID	NO:1	19:								

- - (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 26 base pairs

	(B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
5	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:	
	GACCTCGAGT CCTTCTACAC CTCATC	26
	(121) INFORMATION FOR SEQ ID NO:120:	
10	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
:	(ii) MOLECULE TYPE: DNA (genomic)	•
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:	
15	TGCTCTAGAT TCCAGATAGG TGAAAACTTG	30
	(122) INFORMATION FOR SEQ ID NO:121:	
20	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1416 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (genomic) (xi) SEQUENCE DESCRIPTION: SEQ ID NO:121: 	·.
	ATGGATATTC TTTGTGAAGA AAATACTTCT TTGAGCTCAA CTACGAACTC CCTAATGCAA	60
25	TTAAATGATG ACAACAGGCT CTACAGTAAT GACTTTAACT CCGGAGAAGC TAACACTTCT	120
	GATGCATTTA ACTGGACAGT CGACTCTGAA AATCGAACCA ACCTTTCCTG TGAAGGGTGC	180
	CTCTCACCGT CGTGTCTCTC CTTACTTCAT CTCCAGGAAA AAAACTGGTC TGCTTTACTG	240
	ACAGCCGTAG TGATTATTCT AACTATTGCT GGAAACATAC TCGTCATCAT GGCAGTGTCC	300
	CTAGAGAAAA AGCTGCAGAA TGCCACCAAC TATTTCCTGA TGTCACTTGC CATAGCTGAT	360
30	ATGCTGCTGG GTTTCCTTGT CATGCCCGTG TCCATGTTAA CCATCCTGTA TGGGTACCGG	420
	TGGCCTCTGC CGAGCAAGCT TTGTGCAGTC TGGATTTACC TGGACGTGCT CTTCTCCACG	480
	GCCTCCATCA TGCACCTCTG CGCCATCTCG CTGGACCGCT ACGTCGCCAT CCAGAATCCC	540
	ATCCACCACA GCCGCTTCAA CTCCAGAACT AACCCATTTTC TOAAAATCAT TOCTOTTCA	

WO 00/22129 PCT/US99/23938

05

				93			
	ACCATATCAG	TAGGTATATC	CATGCCAATA	CCAGTCTTTG	GGCTACAGGA	CGATTCGAAG	660
	GTCTTTAAGG	AGGGGAGTTG	CTTACTCGCC	GATGATAACT	TTGTCCTGAT	CGGCTCTTTT	720
	GTGTCATTTT	TCATTCCCTT	AACCATCATG	GTGATCACCT	ACTTTCTAAC	TATCAAGTCA	780
	CTCCAGAAAG	AAGCTACTTT	GTGTGTAAGT	GATCTTGGCA	CACGGGCCAA	ATTAGCTTCT	840
5	TTCAGCTTCC	TCCCTCAGAG	TTCTTTGTCT	TCAGAAAAGC	TCTTCCAGCG	GTCGATCCAT	900
	AGGGAGCCAG	GGTCCTACAC	AGGCAGGAGG	ACTATGCAGT	CCATCAGCAA	TGAGCAAAAG	960
	GCATGCAAGG	TGCTGGGCAT	CGTCTTCTTC	CTGTTTGTGG	TGATGTGGTG	CCCTTTCTTC	1020
	ATCACAAACA	TCATGGCCGT	CATCTGCAAA	GAGTCCTGCA	ATGAGGATGT	CATTGGGGCC	1080
	CTGCTCAATG	TGTTTGTTTG	GATCGGTTAT	CTCTCTTCAG	CAGTCAACCC	ACTAGTCTAC	1140
10	ACACTGTTCA	ACAAGACCTA	TAGGTCAGCC	TTTTCACGGT	ATATTCAGTG	TCAGTACAAG	1200
	GAAAACAAAA	AACCATTGCA	GTTAATTTTA	GTGAACACAA	TACCGGCTTT	GGCCTACAAG	-1260
	TCTAGCCAAC	TTCAAATGGG	ACAAAAAAG	AATTCAAAGC	AAGATGCCAA	GACAACAGAT	1320
	AATGACTGCT	CAATGGTTGC	TCTAGGAAAG	CAGTATTCTG	AAGAGGCTTC	TAAAGACAAT	-1380
	AGCGACGGAG	TGAATGAAAA	GGTGAGCTGT	GTGTGA			1416
15	(123) INFOR	MATION FOR	SEQ ID NO:1	.22 :			
20	(EQUENCE CHA A) LENGTH: B) TYPE: am C) STRANDED D) TOPOLOGY	471 amino a ino acid NESS:	cids			

- 20
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

Met Asp Ile Leu Cys Glu Glu Asn Thr Ser Leu Ser Ser Thr Thr Asn

25 Ser Leu Met Gln Leu Asn Asp Asp Asn Arg Leu Tyr Ser Asn Asp Phe 20

> Asn Ser Gly Glu Ala Asn Thr Ser Asp Ala Phe Asn Trp Thr Val Asp 35 40

Ser Glu Asn Arg Thr Asn Leu Ser Cys Glu Gly Cys Leu Ser Pro Ser 30 55

> Cys Leu Ser Leu Leu His Leu Gln Glu Lys Asn Trp Ser Ala Leu Leu 65 70 75

					85					90					95	
	Met	: Al	a Va	1 Se 10	r Lei O	u Glı	ı Lys	s Ly:	10!	u Gli	n Ası	a Ala	a Th	110		r Phe
5	Leu	Me	11.	r Le	u Āla	a Ile	e Ala	Ası 120	Met	Let	ı Lei	Gly	/ Pho		ı Va	l Met
	Pro	Va:	l Se:	r Mei	t Lev	ı Thr	: Ile 135	Let	туг	Gly	у Туг	140		Pro	Le	ı Pro
10	Ser 145	Lys	E Let	ı Cys	s Ala	150	Trp	Ile	туг	Leu	Asp 155		Let	Phe	: Se	Thr 160
	Ala	Sei	: Ile	e Met	His 165	Leu	Cys	Ala	Ile	Ser 170	Leu	Asp	Arg	Туг	Val	Ala
	Ile	Glr	Ası	180))	His	His	Ser	Arg	Phe	Asn	Ser	Arg	Thr 190		. Ala
15	Phe	Leu	Lys 195	: Ile	: Ile	Ala	Val	Trp 200	Thr	Ile	Ser	Val	Gly 205		Ser	Met
	Pro	Ile 210	Pro	Val	Phe	Gly	Leu 215	Gln	Asp	Asp	Ser	Lys 220	Val	Phe	Lys	Glu
20	Gly 225	Ser	Cys	Leu	Leu	Ala 230	Asp	Asp	Asn	Phe	Val 235	Leu	Ile	Gly	Ser	Phe 240
					245					250		;			255	Leu
0.5				260				•.	265		Leu			270		
25			2/5					280			Phe		285			
		200					295					300				
30	Ser 305					310					315					320
	Ala				325					330					335	
25	Cys			340					345					350		
35	Cys		333				•	360					365			
	Gly ?	ıyr	Leu	Ser	Ser .	Ala '	Val 1	Asn	Pro :	Leu	Val '	Tyr '	Thr :	Leu	Phe	Asn

BNSDOCID WC 0022129A1 IA>

WO 00/22129 PCT/US99/23938

			370					375					380				
		Lys 385	Thr	Tyr	Arg	Ser	Ala 390	Phe	Ser	Arg	Tyr	Ile 395	Gln	Cys	Gln	Tyr	Lys 400
5			Asn	Lys	Lys	Pro 405		Gln	Leu	Ile	Leu 410	Val	Asn	Thr	Ile	Pro 415	Ala
		Leu	Ala	Tyr	Lys 420	Ser	Ser	Gln	Leu	Gln 425	Met	Gly	Gln	Lys	Lys 430	Asn	Ser
		Lys	Gln	Asp 435	Ala	Lys	Thr	Thr	Asp 440	Asn	Asp	Cys	Ser	Met 445	Val	Ala	Leu
10		Gly	Lys 450	Gln	Tyr	Ser	Glu	Glu 455	Ala	Ser	Lys	Asp	Asn 460	Ser	Asp	Gly	Val
		Asn 465	Glu	Lys	Val	Ser	Cys 470	Val									
	(124)	INF	ORMA	ATION	FOF	SEÇ) ID	NO:1	.23:								
15	,	(i)	(A) (B) (C)	LEN	IGTH: PE: r RANDE	27 nucle	base ic a SS: s	ingl	.rs								
20		(ii) MC	LECU	LE 1	YPE:	DNA	(ge	nomi	.c)							
		(xi) SE	QUEN	ICE I	ESCR	PT1	ON:	SEQ	ID N	10:12	:3:					
	GACCT	CGAG	G TI	GCTI	'AAGA	CTG	AAGC	! .									27
	(125)	INF	ORMA	MOIT	FOR	SEC	ID	NO:1	24:								
25		(i)	(A) (B) (C)	LEN TYP	GTH: E: n ANDE	27 ucle DNES	base ic a S: s	ingl	rs								
	•	(ii) MO	LECU	LE T	YPE:	DNA	(ge	nomi	c)							
30		(xi) SE	QUEN	CE D	ESCR	IPTI	ON:	SEQ	ID N	0:12	4:					
	ATTTC:	TAGA	C AT	'ATGT	AGCT	TGT	ACCG	;									27
	(126)	INF	ORMA	TION	FOR	SEQ	ID	NO:1	25:	٠							
35		(i)	(A) (B) (C)	LEN TYP	GTH: E: n ANDE	137 ucle DNES	7 ba ic a S: s	ingl	airs								

- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

	ATGGTGAACC	TGAGGAATGC	GGTGCATTCA	A TTCCTTGTG	ACCTAATTG	G CCTATTGGTT	60
	TGGCAATGTG	ATATTTCTGT	GAGCCCAGTA	GCAGCTATAC	TAACTGACA	TTTCAATACC	120
5	TCCGATGGTG	GACGCTTCAA	ATTCCCAGAC	GGGGTACAA	ACTGGCCAG	C ACTTTCAATC	180
	GTCATCATAA	TAATCATGAC	AATAGGTGGC	AACATCCTTG	TGATCATGG	AGTAAGCATG	240
	GAAAAGAAAC	TGCACAATGC	CACCAATTAC	TTCTTAATGT	CCCTAGCCAT	TGCTGATATG	300
	CTAGTGGGAC	TACTTGTCAT	GCCCCTGTCT	CTCCTGGCAA	TCCTTTATGA	TTATGTCTGG	360
	CCACTACCTA	GATATTTGTG	CCCCGTCTGG	ATTTCTTTAG	ATGTTTTATT	TTCAACAGCG	420
10	TCCATCATGC	ACCTCTGCGC	TATATCGCTG	GATCGGTATG	TAGCAATACG	TAATCCTATT	480
	GAGCATAGCC	GTTTCAATTC	GCGGACTAAG	GCCATCATGA	AGATTGCTAT	TGTTTGGGCA	540
:	ATTTCTATAG	GTGTATCAGT	TCCTATCCCT	GTGATTGGAC	TGAGGGACGA	AGAAAAGGTG	600
	TTCGTGAACA	ACACGACGTG	CGTGCTCAAC	GACCCAAATT	TCGTTCTTAT	TGGGTCCTTC	660
	GTAGCTTTCT	TCATACCGCT	GACGATTATG	GTGATTACGT	ATTGCCTGAC	CATCTACGTT	720
15	CTGCGCCGAC	AAGCTTTGAT	GTTACTGCAC	GGCCACACCG	AGGAACCGCC	TGGACTAAGT	780
	CTGGATTTCC	TGAAGTGCTG	CAAGAGGAAT	ACGGCCGAGG	AAGAGAACTC	TGCAAACCCT	840
	AACCAAGACC	AGAACGCACG	CCGAAGAAAG	AAGAAGGAGA	GACGTCCTAG	GGGCACCATG	900
	CAGGCTATCA	ACAATGAAAG	AAAAGCTTCG	AAAGTCCTTG	GGATTGTTTT	CTTTGTGTTT	960
	CTGATCATGT	GGTGCCCATT	TTTCATTACC	AATATTCTGT	CTGTTCTTTG	TGAGAAGTCC	1020
20	TGTAACCAAA	AGCTCATGGA	AAAGCTTCTG	AATGTGTTTG	TTTGGATTGG	CTATGTTTGT	1080
	TCAGGAATCA	ATCCTCTGGT	GTATACTCTG	TTCAACAAAA	TTTACCGAAG	GGCATTCTCC	1140
	AACTATTTGC	GTTGCAATTA	TAAGGTAGAG	AAAAAGCCTC	CTGTCAGGCA	GATTCCAAGA	1200
	GTTGCCGCCA	CTGCTTTGTC	TGGGAGGGAG	CTTAATGTTA	ACATTTATCG	GCATACCAAT	1260
25						AGTTGAGAAT	1320
25	TTAGAGTTAC	CAGTAAATCC	CTCCAGTGTG	GTTAGCGAAA	GGATTAGCAG	TGTGTGA	1377
	(127) INFOR	MATION FOR	SEO ID NO.1	26.			

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 458 amino acids
 - (B) TYPE: amino acid

99

- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:
- 5 Met Val Asn Leu Arg Asn Ala Val His Ser Phe Leu Val His Leu Ile 10 Gly Leu Leu Val Trp Gln Cys Asp Ile Ser Val Ser Pro Val Ala Ala Ile Val Thr Asp Ile Phe Asn Thr Ser Asp Gly Gly Arg Phe Lys Phe 10 40 Pro Asp Gly Val Gln Asn Trp Pro Ala Leu Ser Ile Val Ile Ile Ile Ile Met Thr Ile Gly Gly Asn Ile Leu Val Ile Met Ala Val Ser Met 70 15 Glu Lys Lys Leu His Asn Ala Thr Asn Tyr Phe Leu Met Ser Leu Ala 90 Ile Ala Asp Met Leu Val Gly Leu Leu Val Met Pro Leu Ser Leu Leu 105 Ala Ile Leu Tyr Asp Tyr Val Trp Pro Leu Pro Arg Tyr Leu Cys Pro 20 115 120 Val Trp Ile Ser Leu Asp Val Leu Phe Ser Thr Ala Ser Ile Met His 135 Leu Cys Ala Ile Ser Leu Asp Arg Tyr Val Ala Ile Arg Asn Pro Ile 155 25 Glu His Ser Arg Phe Asn Ser Arg Thr Lys Ala Ile Met Lys Ile Ala 165 . Ile Val Trp Ala Ile Ser Ile Gly Val Ser Val Pro Ile Pro Val Ile 185 Gly Leu Arg Asp Glu Glu Lys Val Phe Val Asn Asn Thr Thr Cys Val 30 . 200 Leu Asn Asp Pro Asn Phe Val Leu Ile Gly Ser Phe Val Ala Phe Phe Ile Pro Leu Thr Ile Met Val Ile Thr Tyr Cys Leu Thr Ile Tyr Val 230 35 Leu Arg Arg Gln Ala Leu Met Leu Leu His Gly His Thr Glu Glu Pro

250

		Pro	o Gly	/ Let	260	Leu	Asp	Phe	: Leu	Lys 265	. Cys	Cys	Lys	Arg	270		Ala
		Glı	ı Glı	275	ı Asn	Ser	` Ala	Asn	Pro 280	Asn	Gln	Asp	Gln	Asn 285		Arg	Arg
5		Arg	290	Lys	Lys	Glu	Arg	Arg 295	. Pro	Arg	Gly	Thr	Met 300	Gln	Ala	Ile	Asn
		Asn 305	Glu	Arg	Lys	Ala	Ser 310	Lys	Val	Leu	Gly	Ile 315	Val	Phe	Phe	Val	Phe 320
10		Leu	Ile	Met	Trp	Cys 325	Pro	Phe	Phe	Ile	Thr 330	Asn	Ile	Leu	Ser	Val 335	Leu
		Cys	Glu	Lys	Ser 340	Cys	Asn	Gln	Lys	Leu 345	Met	Glu	Lys	Leu	Leu 350	Asn	Val
		Phe	Val	Trp 355	Ile	Gly	Tyr	Val	Cys 360	Ser	Gly	Ile	Asn	Pro 365	Leu	Val	Tyr
15		Thr	Leu 370	Phe	Asn	Lys	Ile	Tyr 375	Arg	Arg	Ala	Phe	Ser 380	Asn	Tyr	Leu	Arg
		Cys 385	Asn	Tyr	Lys	Val	Glu 390	Lys	Lys	Pro	Pro	Val 395	Arg	Gln	Ile	Pro	Arg 400
20		Val	Ala	Ala	Thr	Ala 405	Leu	Ser	Gĺy		Glu 410	Leu	Asn	Val	Asn	Ile 415	Tyr
		Arg	His	Thr	Asn 420	Glu	Pro	Val	Ile	Glu 425	Lys	Ala	Ser	Asp	Asn 430	Glu	Pro
		Gly	Ile	Glu 435	Met	Gln	Val	Glu	Asn 440	Leu	Glu	Leu .		Val 445	Asn	Pro	Ser
25		Ser	Val 450	Val	Ser	Glu		Ile 455	Ser	Ser	Val						
	(128)	INF	FORMA	TION	FOR	SEQ	ID	NO:1	27:								
30		(i)	(A) (B) (C)	LEN TYP STR	E CH GTH: E: n: ANDE: OLOG	30 l ucle: DNES:	base ic a S: s:	pai: cid ingl	rs								
٧		(ii) MO	LECU:	LE T	YPE:	DNA	(gei	nomi	2)			, r	i ne	e e eg	Mitter is	p
35	0000								SEQ]	ID NO	0:127	' :					
دد	GGTAA																30
	(129)	INF	ORMA'	rion	FOR	SEQ	ID 1	NO:12	28:								

_ 5	(A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:	
	TCCGAATTCT CTGTAGACAC AAGGCTTTGG	· 3
	(130) INFORMATION FOR SEQ ID NO:129:	
10	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1068 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
15	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:	
	ATGGATCAGT TCCCTGAATC AGTGACAGAA AACTTTGAGT ACGATGATTT GGCTGAGGCC	· 6
	TGTTATATTG GGGACATCGT GGTCTTTGGG ACTGTGTTCC TGTCCATATT CTACTCCGTC	12
	ATCTTTGCCA TTGGCCTGGT GGGAAATTTG TTGGTAGTGT TTGCCCTCAC CAACAGCAAG	186
20	AAGCCCAAGA GTGTCACCGA CATTTACCTC CTGAACCTGG CCTTGTCTGA TCTGCTGTTT	240
	GTAGCCACTT TGCCCTTCTG GACTCACTAT TTGATAAATG AAAAGGGCCT CCACAATGCC	300
	ATGTGCAAAT TCACTACCGC CTTCTTCTTC ATCGGCTTTT TTGGAAGCAT ATTCTTCATC	360
	ACCGTCATCA GCATTGATAG GTACCTGGCC ATCGTCCTGG CCGCCAACTC CATGAACAAC	420
	CGGACCGTGC AGCATGGCGT CACCATCAGC CTAGGCGTCT GGGCAGCAGC CATTTTGGTG	480
25	GCAGCACCCC AGTTCATGTT CACAAAGCAG AAAGAAAATG AATGCCTTGG TGACTACCCC	540
	GAGGTCCTCC AGGAAATCTG GCCCGTGCTC CGCAATGTGG AAACAAATTT TCTTGGCTTC	600
	CTACTCCCCC TGCTCATTAT GAGTTATTGC TACTTCAGAA TCATCCAGAC GCTGTTTTCC	66Ó
	TGCAAGAACC ACAAGAAAGC CAAAGCCATT AAACTGATCC TTCTGGTGGT CATCGTGTTT	720
30	TTCCTCTTCT GGACACCCTA CAACGTTATG ATTTTCCTGG AGACGCTTAA GCTCTATGAC	780
	TTCTTTCCCA GTTGTGACAT GAGGAAGGAT CTGAGGCTGG CCCTCAGTGT GACTGAGACG	840
	GTTGCATTTA GCCATTGTTG CCTGAATCCT CTCATCTATG CATTTGCTGG GGAGAAGTTC	900
	AGAAGATACC TTTACCACCT GTATGGGAAA TGCCTGGCTG TCCTGTGTGG GCGCTCAGTC	960

CACGTTGATT TCTCCTCATC TGAATCACAA AGGAGCAGGC ATGGAAGTGT TCTGAGCAGC 1020

				•													
	AATTT	TACI	T AC	CAC	ACGAC	TGP	ATGGI	AGAT	GCAT	TGC	rcc 1	TCTC	CTGA				1068
	(131)	INF	ORMA	TION	1 FOF	SEÇ	Q ID	NO: 3	130:								
5		(i)	(A) (B) (C)	LEN TYI STI	CE CH NGTH: PE: & RANDE	355 mino EDNES	ami aci SS:	no a	acids		:	· .					
		(ii	.) MC	LECU	JLE T	TYPE:	pro	oteir	1					,	•		
10		(xi) SI	EQUE1	NCE I	DESC	RIPTI	ON:	SEQ	ID 1	10:13	30:					
		Met 1	Asp	Gln	Phe	Pro 5	Glu	Ser	Val	Thr	Glu 10	Asn	Phe	Glu	Tyr	Asp 15	Asp
		Leu	Ala	Glu	Ala 20	Cys	Tyr	Ile	Gly	Asp 25	Ile	Val	Val	Phe	Gly 30	Thr	Val
15		Phe	Leu	Ser 35	Ile	Phe	Tyr	Ser	Val 40	Ile	Phe	Ala	Ile	Gly 45	Leu	Val	Gly
		Asn	Leu 50	Leu	Val	Val	Phe	Ala 55	Leu	Thr	Asn	Ser	Lys 60	Lys	Pro	Lys	Ser
20		Val 65	Thr	Asp	Ile	Tyr	Leu 70	Leu	Asn	Leu	Ala	Leu 75	Ser	Asp	Leu	Leu	Phe 80
		Val	Ala	Thr	Leu	Pro 85	Phe	Trp	Thr	His	Tyr 90	Leu	Ile	Asn		Lys 95	Gly
		Leu	His	Asn	Ala 100	Met	Cys	Lys	Phe	Thr 105	Thr	Ala	Phe	Phe	Phe 110	Ile	Gly
25		Phe	Phe	Gly 115	Ser	Ile	Phe	Phe	11e 120	Thr	Val	Ile	Ser	Ile 125		Arg	Tyr
		Leu	Ala 130	Ile	Val	Leu	Ala	Ala 135	Asn	Ser	Met	Asn	Asn 140	Arg	Thr	Val	Gln
30		His 145	Gly	Val	Thr	Ile	Ser 150	Leu	Gly	Val	Trp	Ala 155	Ala	Ala	Ile	Leu	Val 160
		Ala	Ala	Pro	Gln	Phe 165	Met	Phe	Thr	Lys	Gln 170	Lys	Glu	Asn	Glu	Cys 175	Leu
		Gly	Asp	Tyr	Pro 180	Glu	Val	Leu	Gln	Glu 185	Ile	Trp	Pro	Val	Leu 190	Arg	Asn
35		Val	Glu	Thr 195	Asn	Phe	Leu	Gly	Phe 200	Leu	Leu	Pro	Leu	Leu 205	Ile	Met	Ser

	Tyr	Cys 210		Phe	Arg	Ile	Ile 215	Gln	Thr	Leu	Phe	Ser 220	Cys	Lys	Asn	His	
	Lys 225	Lys	Ala	Lys	Ala	Ile 230	Lys	Leu	Ile	Leu	Leu 235	Val	Val	Ile	Val	Phe 240	
5	Phe	Leu	Phe	Trp	Thr 245	Pro	Tyr	Asn	Val	Met 250	Ile	Phe	Leu	Glu	Thr 255	Leu	
	Lys	Leu	Tyr	Asp 260	Phe	Phe	Pro	Ser	Cys 265	Asp	Met	Arg	Lys	Asp 270	Leu	Arg	
10	Leu	Ala	Leu 275	Ser	Val	Thr	Glu	Thr 280	Val	Ala	Phe	Ser	His 285	Cys	Cys	Leu	
	Asn	Pro 290	Leu	Ile	Tyr	Ala	Phe 295	Ala	Gly	Glu	Lys	Phe 300	Arg	Arg	Tyr	Leu	
	Туг 305	His	Leu	Tyr	Gly	Lys 310	Cys	Leu	Ala	Val	Leu 315	Cys	Gly	Arg	Ser	Val 320	
15	His	Val	Asp	Phe	Ser 325	Ser	Ser	Glu	Ser	Gln 330	_	Ser	Arg	His	Gly 335	Ser	. •
	Val	Leu	Ser	Ser 340	Asn	Phe	Thr	Tyr	His 345	Thr	Ser	Asp	Gly	Asp 350	Ala	Leu	
20	Leu	Leu	Leu 355			-									•.		- r'
	(132) IN	FORM	ATIO1	1 FOF	SEC) ID	NO:1	131:							•		,
25	(i	(B)	LEN TYPE	IGTH: PE: r	32 nucle DNES	base ic a SS: s	e pai acid singl	rs									
	(i	i) MC						nomi	C)								
		i) SF					_			IO:13	1.						
	GATCTCCA																32
30	(133) IN																-
35) SE((A) (B) (C)	QUENC LEN TYP STR	E CH IGTH: PE: n	ARAC 30 ucle	TERI base ic a	STIC pai cid ingl	:s: .rs									
	(i:	i) MC	DLECU	LE T	YPE:	DNA	(ge	nomi	c)								
	(x:	i) SE	QUEN	CE D	ESCR	IPTI	ON:	SEQ	ID N	0:13	2:						

	CTCCTTCGGT CCTCCTATCG TTGTCAGAAG	30
	(134) INFORMATION FOR SEQ ID NO:133:	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:	
10	AGAAGGCCAA GATCGCGCGG CTGGCCCTCA	30
	(135) INFORMATION FOR SEQ ID NO:134:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:134:	
	CGGCGCCACC GCACGAAAAA GCTCATCTTC	30
20	(136) INFORMATION FOR SEQ ID NO:135: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 base pairs (B) TYPE: nucleic acid	
25	(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:135:	
	GCCAAGAAGC GGGTGAAGTT CCTGGTGGTG GCA	33
	(137) INFORMATION FOR SEQ ID NO:136:	
30	(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 30 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
35	(ii) MOLECULE TYPE: DNA (genomic)	

	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:	
	CAGGCGGAAG GTGAAAGTCC TGGTCCTCGT	3
	(138) INFORMATION FOR SEQ ID NO:137:	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
10	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:	
	CGGCGCCTGC GGGCCAAGCG GCTGGTGGTG GTG	33
	(139) INFORMATION FOR SEQ ID NO:138:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:138:	
20	CCAAGCACAA AGCCAAGAAA GTGACCATCA C	31
	(140) INFORMATION FOR SEQ ID NO:139:	÷
25	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:139:	
	GCGCCGGCGC ACCAAATGCT TGCTGGTGGT	30
30	(141) INFORMATION FOR SEQ ID NO:140:	
35	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 41 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	

	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:	
	CAAAAAGCTG AAGAAATCTA AGAAGATCAT CTTTATTGTC G	4:
	(142) INFORMATION FOR SEQ ID NO:141:	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
10	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:141:	
	CAAGACCAAG GCAAAACGCA TGATCGCCAT	3 (
	(143) INFORMATION FOR SEQ ID NO:142:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:142:	
	GTCAAGGAGA AGTCCAAAAG GATCATCATC	30
	(144) INFORMATION FOR SEQ ID NO:143:	
25	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:143:	
30	CGCCGCGTGC GGGCCAAGCA GCTCCTGCTC	3 (
	(145) INFORMATION FOR SEQ ID NO:144:	
35	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	

THE PARTY

	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:144:	
	CCTGATAAGC GCTATAAAAT GGTCCTGTTT CGA	33
	(146) INFORMATION FOR SEQ ID NO:145:	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 36 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
10	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:145:	
	GAAAGACAAA AGAGAGTCAA GAGGATGTCT TTATTG	36
	(147) INFORMATION FOR SEQ ID NO:146:	. .
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	:
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:	
	CGGAGAAAGA GGGTGAAACG CACAGCCATC GCC	33
	(148) INFORMATION FOR SEQ ID NO:147:	
25	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:147:	
30	AAGCTTCAGC GGGCCAAGGC ACTGGTCACC	30 [:]
	(149) INFORMATION FOR SEQ ID NO:148:	
35	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	

	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:148:	
	CAGCGGCAGA AGGCAAAAG GGTGGCCATC	30
	(150) INFORMATION FOR SEQ ID NO:149:	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
10	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:149:	
	CGGCAGAAGC CGAAGCGCAT GATCCTCGCG	30
	(151) INFORMATION FOR SEQ ID NO:150:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:150:	
	GAGCGCAACA AGGCCAAAAA GGTGATCATC	30
	(152) INFORMATION FOR SEQ ID NO:151:	
25	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:151:	
30.	GGTGTAAACA AAAAGGCTAA AAACACAATT ATTCTTATT	3 9
	(153) INFORMATION FOR SEQ ID NO:152:	
35	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 27 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	

	(ii) MOLECULE TYPE: DNA (genomic)	
	. (xi) SEQUENCE DESCRIPTION: SEQ ID NO:152:	
•	GAGAGCCAGC TCAAGAGCAC CGTGGTG	27
	(154) INFORMATION FOR SEQ ID NO:153:	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
10	(ii) MOLECULE TYPE: DNA (genomic)	
÷	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:153:	
	CCACAAGCAA ACCAAGAAAA TGCTGGCTGT	30
	(155) INFORMATION FOR SEQ ID NO:154:	•
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	•
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:154:	÷
	CATCAAGTGT ATCATGTGCC AAGTACGCCC	30
	(156) INFORMATION FOR SEQ ID NO:155:	
25	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 34 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:155:	
30	CTAGAGAGTC AGATGAAGTG TACAGTAGTG GCAC	34
	(157) INFORMATION FOR SEQ ID NO:156:	
35	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 36 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	

	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:156:	
	CGGACAAAAG TGAAAACTAA AAAGATGTTC CTCATT	16
	(158) INFORMATION FOR SEQ ID NO:157:	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
10	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:157:	
2	GCTGAGGTTC GCAATAAACT AACCATGTTT GTG	3
	(159) INFORMATION FOR SEQ ID NO:158:	
15	(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 29 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:158:	
	GGGAGGCCGA GCTGAAAGCC ACCCTGCTC	9
	(160) INFORMATION FOR SEQ ID NO:159:	
25	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:159:	
30	CAAGATCAAG AGAGCCAAAA CCTTCATCAT G	31
	(161) INFORMATION FOR SEQ ID NO:160:	
35	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	

WO 00/22129 PCT/US99/23938

' 111

	(ii) MOLECULE TYPE: DNA (genomic)
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:160:
r	CCGGAGACAA GTGAAGAAGA TGCTGTTTGT C
	(162) INFORMATION FOR SEQ ID NO:161:
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear
10	(ii) MOLECULE TYPE: DNA (genomic)
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:161:
٠٠.	GCAAGGACCA GATCAAGCGG CTGGTGCTCA
	(163) INFORMATION FOR SEQ ID NO:162:
15	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 34 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear
	(ii) MOLECULE TYPE: DNA (genomic)
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:162:
	CAAGAAAGCC AAAGCCAAGA AACTGATCCT TCTG
	(164) INFORMATION FOR SEQ ID NO:163:
25	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1068 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear
	(ii) MOLECULE TYPE: DNA (genomic)
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:163:
30	ATGGAAGATT TGGAGGAAAC ATTATTTGAA GAATTTGAAA ACTATTCCTA TGACCTAGAC 60
	TATTACTCTC TGGAGTCTGA TTTGGAGGAG AAAGTCCAGC TGGGAGTTGT TCACTGGGTC 120
	TCCCTGGTGT TATATTGTTT GGCTTTTGTT CTGGGAATTC CAGGAAATGC CATCGTCATT 180
	TGGTTCACGG GGCTCAAGTG GAAGAAGACA GTCACCACTC TGTGGTTCCT CAATCTAGCC 240
	ATTGCGGATT TCATTTTCT TCTCTTTCTG CCCCTGTACA TCTCCTATGT GGCCATGAAT 300

	TTCCACTGGC CCTTTGGCAT CTGGCTGTGC AAAGCCAATT CCTTCACTGC CCAGTTGA	AC 360
	ATGTTTGCCA GTGTTTTTT CCTGACAGTG ATCAGCCTGG ACCACTATAT CCACTTGA	TC 420
	CATCCTGTCT TATCTCATCG GCATCGAACC CTCAAGAACT CTCTGATTGT CATTATAT	TC 480
	ATCTGGCTTT TGGCTTCTCT AATTGGCGGT CCTGCCCTGT ACTTCCGGGA CACTGTGG	AG 540
5	TTCAATAATC ATACTCTTTG CTATAACAAT TTTCAGAAGC ATGATCCTGA CCTCACTT	TG 600
	ATCAGGCACC ATGTTCTGAC TTGGGTGAAA TTTATCATTG GCTATCTCTT CCCTTTGC	TA 660
	ACAATGAGTA TTTGCTACTT GTGTCTCATC TTCAAGGTGA AGAAGCGAAC AGTCCTGA	TC 720
	TCCAGTAGGC ATAAGTGGAC AATTCTGGTT GTGGTTGTGG CCTTTGTGGT TTGCTGGA	.CT 780
	CCTTATCACC TGTTTAGCAT TTGGGAGCTC ACCATTCACC ACAATAGCTA TTCCCACC	AT 840
10	GTGATGCAGG CTGGAATCCC CCTCTCCACT GGTTTGGCAT TCCTCAATAG TTGCTTGA	AC 900
	CCCATCCTTT ATGTCCTAAT TAGTAAGAAG TTCCAAGCTC GCTTCCGGTC CTCAGTTG	CT 960
	GAGATACTCA AGTACACACT GTGGGAAGTC AGCTGTTCTG GCACAGTGAG TGAACAGC	TC 1020
	AGGAACTCAG AAACCAAGAA TCTGTGTCTC CTGGAAACAG CTCAATAA	1068
	(165) INFORMATION FOR SEQ ID NO:164:	•
15	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 355 amino acids	
	(B) TYPE: amino acid (C) STRANDEDNESS:	
	(D) TOPOLOGY: not relevant	
20	(ii) MOLECULE TYPE: protein	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:164:	:
	Met Glu Asp Leu Glu Glu Thr Leu Phe Glu Glu Phe Glu Asn Ty 1 5 10 15	
25	Tyr Asp Leu Asp Tyr Tyr Ser Leu Glu Ser Asp Leu Glu Glu Ly 20	s Val
	Gln Leu Gly Val Val His Trp Val Ser Leu Val Leu Tyr Cys Le 35 40 45	u Ala
	Phe Val Leu Gly Ile Pro Gly Asn Ala Ile Val Ile Trp Phe Th 50 55 60	r Gly
30	Leu Lys Trp Lys Lys Thr Val Thr Thr Leu Trp Phe Leu Asn Le	u Ala 80
	Ile Ala Asp Phe Ile Phe Leu Leu Phe Leu Pro Leu Tyr Ile Se 85 90 95	-

		Val	Ala	Met	Asn	Phe	His	Trp	Pro	Phe 105	Glý	Ile	Trp	Leu	Cys 110	Lys	Ala
		Asn	Ser	Phe 115	Thr	Ala	Gln	Leu	Asn 120	Met	Phe	Ala	Ser	Val 125	Phe	Phe	Leu
5		Thr	Val 130	Ile	Ser	Leu	Asp	His 135	Tyr	Ile	His	Leu	Ile 140	His	Pro	Val	Leu
		Ser 145	His	Arg	His	Arg	Thr 150	Leu	Lys	Asn	Ser	Leu 155	Ile	Val	Ile	Ile	Phe 160
10		Ile	Trp	Leu	Leu	Ala 165	Ser	Leu	Ile	Gly	Gly 170	Pro	Ala	Leu	Tyr	Phe 175	Arg
		Asp	Thr	Val	Glu 180	Phe	Asn	Asn	His	Thr 185	Leu	Cys	Tyr	Asn	Asn 190	Phe	Gln
	;	Lys	His	Asp 195	Pro	Asp	Leu	Thr	Leu 200	Ile	Arg	His	His	Val 205	Leu	Thr	Trp
15		Val	Lys 210	Phe	Ile	Ile	Gly	Tyr 215	Leu	Phe	Pro	Leu	Leu 220	Thr	Met	Ser	Ile
		Cys 225	Tyr	Leu	Cys	Leu	Ile 230	Phe	Lys	Val	Lys	Lys 235	Arg	Thr	Val	Leu	Ile 240
20		Ser	Ser	Arg	His	Lys 245	Trp	Thr	Ile	Leu	Val 250	Val	Val	Val	Ala	Phe 255	Val
		Val	Cys	Trp	Thr 260	Pro	Tyr	His	Leu	Phe 265	Ser	Ile	Trp	Glu	Leu 270	Thr	Ile
		His	His	Asn 275	Ser	Tyr	Ser	His	His 280	Val	Met	Gln	Ala	Gly 285	Ile	Pro	Leu
25		Ser	Thr 290	Gly	Leu	Ala	Phe	Leu 295	Asn	Ser	Cys	Leu	Asn 300	Pro	Ile	Leu	Tyr
		Val 305	Leu	Ile	Ser	Lys	Lys 310	Phe	Gln	Ala	Arg	Phe 315	Arg	Ser	Ser	Val	Ala 320
30		Glu	Ile	Leu	Lys	Tyr 325	Thr	Leu	Trp	Glu	Val 330	Ser	Cys	Ser	Gly	Thr 335	Val
		Ser	Glu	Gln	Leu 340	Arg	Asn	Ser	Glu	Thr 345	Lys	Asn	Leu	Cys	Leu 350	Leu	Glu
		Thr	Ala	Gln 355													

- 35 (166) INFORMATION FOR SEQ ID NO:165:
 - (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1089 base pairs

(B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:165:

5

- ATGGGCAACC ACACGTGGGA GGGCTGCCAC GTGGACTCGC GCGTGGACCA CCTCTTTCCG 60 CCATCCCTCT ACATCTTTGT CATCGGCGTG GGGCTGCCCA CCAACTGCCT GGCTCTGTGG GCGGCCTACC GCCAGGTGCA ACAGCGCAAC GAGCTGGGCG TCTACCTGAT GAACCTCAGC 180 ATCGCCGACC TGCTGTACAT CTGCACGCTG CCGCTGTGGG TGGACTACTT CCTGCACCAC 240 10 GACAACTGGA TCCACGGCCC CGGGTCCTGC AAGCTCTTTG GGTTCATCTT CTACACCAAT 300 ATCTACATCA GCATCGCCTT CCTGTGCTGC ATCTCGGTGG ACCGCTACCT GGCTGTGGCC 360 CACCCACTCC GCTTCGCCCG CCTGCGCCGC GTCAAGACCG CCGTGGCCGT GAGCTCCGTG 420 GTCTGGGCCA CGGAGCTGGG CGCCAACTCG GCGCCCCTGT TCCATGACGA GCTCTTCCGA 480 GACCGCTACA ACCACACCTT CTGCTTTGAG AAGTTCCCCA TGGAAGGCTG GGTGGCCTGG 540 15 ATGAACCTCT ATCGGGTGTT CGTGGGCTTC CTCTTCCCGT GGGCGCTCAT GCTGCTGTCG 600 TACCGGGGCA TCCTGCGGGC CGTGCGGGGC AGCGTGTCCA CCGAGCGCCA GGAGAAGGCC 660 AAGATCGCGC GGCTGGCCCT CAGCCTCATC GCCATCGTGC TGGTCTGCTT TGCGCCCTAT 720 CACGTGCTCT TGCTGTCCCG CAGCGCCATC TACCTGGGCC GCCCCTGGGA CTGCGGCTTC 780 GAGGAGCGCG TCTTTTCTGC ATACCACAGC TCACTGGCTT TCACCAGCCT CAACTGTGTG 840 20 GCGGACCCCA TCCTCTACTG CCTGGTCAAC GAGGGCGCCC GCAGCGATGT GGCCAAGGCC 900 CTGCACAACC TGCTCCGCTT TCTGGCCAGC GACAAGCCCC AGGAGATGGC CAATGCCTCG 960 CTCACCCTGG AGACCCCACT CACCTCCAAG AGGAACAGCA CAGCCAAAGC CATGACTGGC 1020 AGCTGGGCGG CCACTCCGCC TTCCCAGGGG GACCAGGTGC AGCTGAAGAT GCTGCCGCCA 1080 GCACAATGA 1089
- 25 (167) INFORMATION FOR SEQ ID NO:166:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 362 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
- 30 (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein

•								-		•						
	(xi) SE	QUEN	CE D	ESCR	IPTI	ON:	SEQ	ID N	10:16	6:					
	Met 1	Gly	Asn	His	Thr 5	Trp	Glu	Gly	Cys	His 10	Val	Asp	Ser	Arg	Val 15	Asp
5	His	Leu	Phe	Pro 20	Pro	Ser	Leu	Tyr	Ile 25	Phe	Val	Ile	Gly	Val 30	Gly	Leu
•	Pro	Thr	Asn 35	Cys	Leu	Ala	Leu	Trp 40	Ala	Ala	Tyr	Arg	Gln 45	Val	Gln	Gln
•	Arg	Asn 50	Glu	Leu	Gly	Val	Tyr 55	Leu	Met	Asn	Leu	Ser 60	Ile	Ala	Asp	Leu
10	Leu 65	Tyr	Ile	Cys	Thr	Leu 70	Pro	Leu	Trp	Val	Asp 75	Tyr	Phe	Leu	His	His 80
	Asp	Asn	Trp	Ile	His 85	Gly	Pro	Gly	Ser	Cys 90	Lys	Leu	Phe	Gly	Phe 95	Ile
15	Phe	Tyr	Thr	Asn 100	Ile	Tyr	Ile	Ser	Ile 105	Ala	Phe	Leu	Cys	Cys 110	Ile	Ser .'
	Val	Asp	Arg 115	Tyr	Leu	Ala	Val	Ala 120	His	Pro	Leu	Arg	Phe 125	Ala	Arg	Leu
	Arg	Arg 130	Val	Lys	Thr	Ala	Val 135	Ala	Val	Ser	Ser	Val 140		Trp	Ala	Thr
20	Glu 145	Leu	Gly	Ala	Asn	Ser 150		Pro	Leu	Phe	His 155	Asp	Glu	Leu	Phe	Arg 160
	Asp	Arg	Tyr	Asn	His 165		Phe	Cys	Phe	Glu 170		Phe	Pro	Met	Glu 175	Gly
25	Trp	Val	Ala	Trp 180	Met	Asn	Leu	Tyr	Arg 185		Phe	Val	Gly	Phe 190		Phe
	Pro	Trp	Ala 195		Met	Leu	Leu	Ser 200		Arg	Gly	Ile	Leu 205	Arg	Ala	Val
	Arg	Gly 210		Val	Ser	Thr	Glu 215		Gln	Glu	Lys	Ala 220		Ile	Ala	. Arg
30	Leu 225		Leu	Ser	Leu	11e 230		Ile	· Val	. Leu	Val 235		: Phe	Ala	Pro	Tyr 240
	His	: Val	. Leu	. Leu	Leu 245		Arg	Ser	Ala	11e 250		Lev	ı Gly	/ Arg	255	Trp
35	Ası	Cys	s Gly	Phe 260		Glu	a Arg	(Val	. Phe 265		: Ala	туг	His	Ser 270		Leu

Ala Phe Thr Ser Leu Asn Cys Val Ala Asp Pro Ile Leu Tyr Cys Leu

		275		280		285	
		Asn Glu Gly 290	Ala Arg Ser 295		la Lys Ala 300	Leu His Asn	Leu
5	Leu <i>1</i> 305	Arg Phe Leu	Ala Ser Asp 310			Ala Asn Ala	
	Leu T	Thr Leu Glu	Thr Pro Leu 325	Thr Ser Ly		Ser Thr Ala 335	Lys
	Ala M	Met Thr Gly 340	Ser Trp Ala	Ala Thr Pr 345	o Pro Ser	Gln Gly Asp 350	Gln
10	Val G	Sln Leu Lys 355	Met Leu Pro	Pro Ala Gl 360	n		
	(168) INFO	RMATION FOR	SEQ ID NO:	167:			
15	(i)	(A) LENGTH: (B) TYPE: n	ARACTERISTIC 1002 base pucleic acid DNESS: singl Y: linear	pairs		* *	
	· (ii)	MOLECULE T	YPE: DNA (ge	enomic)		•	
	(xi)	SEQUENCE D	ESCRIPTION:	SEQ ID NO:	167:		
20	ATGGAGTCCT	CAGGCAACCC	AGAGAGCACC	ACCTTTTTT	ACTATGACC:	r tcagagccag	60
	CCGTGTGAGA	ACCAGGCCTG	GGTCTTTGCT	ACCCTCGCCA	CCACTGTCC	T GTACTGCCTG	120
	GTGTTTCTCC	TCAGCCTAGT	GGGCAACAGC	CTGGTCCTGT	GGGTCCTGGT	r gaagtatgag	180
	AGCCTGGAGT	CCCTCACCAA	CATCTTCATC	CTCAACCTGT	GCCTCTCAG	A CCTGGTGTTC	240
	GCCTGCTTGT	TGCCTGTGTG	GATCTCCCCA	TACCACTGGG	GCTGGGTGCT	GGGAGACTTC	300
25	CTCTGCAAAC	TCCTCAATAT	GATCTTCTCC	ATCAGCCTCT	ACAGCAGCAT	CTTCTTCCTG	360
	ACCATCATGA	CCATCCACCG	CTACCTGTCG	GTAGTGAGCC	CCCTCTCCAC	CCTGCGCGTC	420
	CCCACCCTCC	GCTGCCGGGT	GCTGGTGACC	ATGGCTGTGT	GGGTAGCCAG	CATCCTGTCC	480
	TCCATCCTCG	ACACCATCTT	CCACAAGGTG	CTTTCTTCGG	GCTGTGATTA	TTCCGAACTC	540
	ACGTGGTACC	TCACCTCCGT	CTACCAGCAC	AACCTCTTCT	TCCTGCTGTC	CCTGGGGATT	600
30	ATCCTGTTCT	GCTACGTGGA	GATCCTCAGG	ACCCTGTTCC	GCTCACGCTC	CAAGCGGCGC	660
	CACCGCACGA	AAAAGCTCAT	CTTCGCCATC	GTGGTGGCCT	ACTTCCTCAG	CTGGGGTCCC	720
	TACAACTTCA	CCCTGTTTCT	GCAGACGCTG	TTTCGGACCC	AGATCATCCG	GAGCTGCGAG	780

WO 00/22129 PCT/US99/23938

GCCA	AACAG	C AC	CTA	SAAT	A CGC	CCTC	CTC	ATC	rgcc	GCA A	ACCTO	CGCC:	TT C	rccc	ACTG	2 4	340
TGCTI)AAT1	CC -C(GTG	CTCT	A TGI	CTTC	CGTG	GGG	STCA	AGT :	rccg	CACA	CA C	CTGA	AACA?	r !	900
GTTCT	rccgo	SC AC	TTC:	rggt:	r cro	CCGC	CTG	CAG	GCAC(CCA (3CCC2	AGCC	rc G	ATCC	CCA	2 9	960
TCCCC	TGGT	rg co	CTTC	GCCT2	A TGA	AGGGC	CGCC	TCC	r t ct2	ACT (3A			٠		10	002
(169)		SE((A) (B)	QUENC LEI	CE CI NGTH PE: 8	ARAC : 333 amino EDNES	CTERI ami	STIC	CS:	5			•					
		(D)	TO	POLO	GY: r	ot i	rele	vant									
	(13	E) MC	OLECT	JLE :	TYPE:	pro	oteir	מ									
	(xi	i) SI	EQUE	NCE I	DESCI	RIPTI	ON:	SEQ	ID 1	NO:16	58:						
	Met 1	Glu	Ser	Ser	Gly 5	Asn	Pro	Glu	Ser	Thr 10	Thr	Phe	Phe	Tyr	Tyr 15	Asp	•
	Leu	Gln	Ser	Gln 20	Pro	Cys	Glu	Asn	Gln 25	Ala	Trp	Val	Phe	Ala 30	Thr	Leu	•
	Ala	Thr	Thr 35	Val	Leu	Tyr	Cys	Leu 40	Val	Phe	Leu	Leu	Ser 45	Leu	Val	Gly	2
	Asn	Ser 50	Leu	Val	Leu	Trp	Val 55	Leu	Val	Lys	Tyr	Glu 60	Ser	Leu	Glu	Ser	
	Leu 65	Thr	Asn	Ile	Phe	Ile 70	Leu	Asn	Leu	Cys	Leu 75	Ser	Asp	Leu ·	Val	Phe 80	
	Ala	Cys	Leu	Leu	Pro 85	Val	Trp	Ile	Ser	Pro 90	Tyr	His	Trp	Gly	Trp 95	Val	
	Leu	Gly	Asp	Phe 100		Cys	-					Ile	Phe	Ser 110	Ile	Ser	
	Leu	Tyr	Ser 115	Ser	Ile	Phe	Phe	Leu 120	Thr	Ile	Met	Thr	Ile 125	His	Arg	Tyr	
	Leu	Ser 130	Val	Val	Ser	Pro	Leu 135	Ser	Thr	Leu	Arg	Val 140	Pro	Thr	Leu	Arg	
	Cys 145	Arg	Val	Leu	Val	Thr 150	Met	Ala	Val	Trp	Val 155	Ala	Ser	Ile	Leu	Ser 160	
,	Ser	Ile		Asp	Thr 165	Ile	Phe	His	Lys	Val 170	Leu	Ser	Ser	Gly	Cys 175	Asp	
•	Tyr	Ser	Glu	Leu	Thr	Trp	Tyr	Leu	Thr	Ser	Val	Tyr	Gln	His	Asn	Leu	

	Phe		Leu 195	Leu	Ser	Leu	Gly	Ile 200	Ile	Leu	Phe	Cys	Tyr 205	Val	Glu	Ile	
	Leu	Arg 210	Thr	Leu	Phe	Arg	Ser 215	Arg	Ser	Lys	Arg	Arg 220	His	Arg	Thr	Lys	
5	Lys 225	Leu	Ile	Phe	Ala	Ile 230	Val	Val	Ala	Tyr	Phe 235	Leu	Ser	Trp	Gly	Pro 240	
	Tyr	Asn	Phe	Thr	Leu 245	Phe	Leu	Gln	Thr	Leu 250	Phe	Arg	Thr	Gln	Ile 255	Ile	
10	Arg	Ser	Cys	Glu 260	Ala	Lys	Gln	Gln	Leu 265	Glu	Tyr	Ala	Leu	Leu 270	Ile	Cys	
	Arg	Asn	Leu 275	Ala	Phe	Ser	His	Cys 280	Cys	Phe	Asn	Pro	Val 285	Leu	Tyr	Val	
	Phe	Val 290	Gly	Val	Lys	Phe	Arg 295	Thr	His	Leu	Lys	His 300	Val	Leu	Arg	Gln	
15	Phe 305	Trp	Phe	Cys	Arg	Leu 310	Gln	Ala	Pro	Ser	Pro 315	Ala	Ser	Ile	Pro	His 320	
	Ser	Pro	Gly	Ala	Phe 325	Ala	Tyr	Glu	Gly	Ala 330		Phe	Tyr				
	(170) INF	FORM	ATIO	N FO	R SE	Q ID	NO:	169:									
20	(i)	(A) (B) (C)) LE) TY) ST	CE C NGTH PE: RAND	: 98 nucl EDNE	7 ba eic SS:	se p acid sing	airs									
25	(1)			ULE				enom	ic)								
		•		NCE	•					NO:1	69:	•		•	. •		
	ATGGACAA	CG- C	CTCG	TTCT	C GG	AGCC	CTGG	CCC	GCCA	ACG	CATC	GGGC	CC G	GACC	CGGC	G	60
	CTGAGCTG	CT C	CAAC	GCGT	'C GA	CTCI	GGCG	CCG	CTGC	CGG	CGCC	GCTG	GC G	GTGG	CTGT	A	120
	CCAGTTGT	CT A	.CGCG	GTGA	T CI	'GCGC	CGTG	GGI	CTGG	CGG	GCAA	CTCC	GC C	GTGC	TGTA	C	180
30	GTGTTGCT	GC G	GGCG	CCCC	G CA	TGAP	AGACO	GTC	ACCA	ACC	TGTI	CATO	CT C	CAACC	TGGC	C	240
	ATCGCCGA	CG A	GCTC	CTTCA	C GC	TGGT	rgcTo	CCC	ATCA	ACA	TCGC	CGAC	TT (CTGC	TGCG	G	300
-	CAGTGGCC	CT T	'CGGG	GAGC	T CA	ATGTO	CAAG	CTC	ATC	TGG	CTAT	CCGAC	CA (TACA	ACAC	:C	360
	TTCTCCAG	CC I	CTAC	CTTCC	T C	ACCG7	CATO	ag(CGCCG	SACC	GCTA	ACCT	GT (GTGI	TGGC	:C	42
	ACTGCGGA	GT C	:GCGC	CCGGG	T GO	CCG	3CCG(acc	TAC	AGCG	CCG	CGCGC	CGC (GTG	AGCCI	`G	48

	GCCGTGTG	GG G	GATC	GTCAC	C AC	rcgr	CGTG	CTG	CCT	rcg	CAGT	CTTC	GC C	CGGC'	TAGA	C	540
	GACGAGCA	GG G	CCGG	CGCCZ	A GT	CGT	GCTA	GTC:	rttc	CGC .	AGCC	CGAG	GC C	TTCT	GGTG	3	600
	CGCGCGAG	CC G	CCTC	TACAC	G GC	rcgto	GCT G	·GGC	rtcg	CCA	TCCC	CGTG'	rc c	ACCA'	TCTG:	r	660
	GTCCTCTA	TA C	CACC	CTGCI	GT(3CCG(GCTG	CATO	GCA.	rgc	GGCT	GAC:	AG C	CACG	CCAA	3	720
5	GCCCTGGA	GC G	CGCCI	A AGA <i>I</i>	A GC	GGT	GAAG	TTC	CTGG	rgg	TGGC	AATC	CT G	GCGG'	TGTG	C '	780
	СТССТСТС	CT G	GACG	CCCT	A CC	ACCTY	GAGC	ACC	TGG:	rgg	CGCT	CACC	AC C	GACC'	rccc	3	840
	CAGACGCC	GC T	GGTC	ATCGC	TA	rctco	CTAC	TTC	ATCA	CCA	GCCT	GACG'	ra c	GCCA	ACAG	2 :	900
	TGCCTCAA	CC C	CTTC	CTCT	A CG	CCTT	CCTG	GAC	CCA	SCT	TCCG	CAGG	AA C	CTCC	GCCA	3	960
	CTGATAAC	TT G	CCGC	GCGGC	C AGO	CCTG	A.									:	987
10	(171) IN	FORM	ATIO	v FOF	SE(Q ID	NO:	170:				•					
	£)	(B) LEI	CE CH NGTH: PE: & RANDE	: 328 amino	am:	ino a		5								
15		(D) TO	POLOG	3Y: 1	not i	rele	vant	•								•
	(i	.i) M	OLEC	JLE 1	YPE	: pro	otei	n.					•				.*
	(>	i) S	EQUEI	NCE I	DESCI	RIP T	ION:	SEQ	ID 1	10:1	70:						
	Met 1	. Asp	Asn	Ala	Ser 5	Phe	Ser	Glu	Pro	Trp 10	Pro	Ala	Asn	Ala	Ser 15	Gly	
20	Pro	Asp	Pro	Ala 20	Leu	Ser	Cys	Ser	Asn 25	Ala	Ser	Thr	Leu	Ala 30	Pro	Leu	
	Pro	Ala	Pro 35	Leu	Ala	Val	Ala	Val 40	Pro	Val	Val	Tyr	Ala 45	Val	Ile	Cys	
25	Ala	Val 50	Gly	Leu	Ala	Gly	Asn 55	Ser	Ala	Val	Leu	Tyr 60	Val	Leu	Leu	Arg	
	Ala 65	Pro	Arg	Met	Lys	Thr 70	Val	Thr	Asn	Leu	Phe 75	Ile	Leu	Asn	Leu	Ala 80	
	Ile	e Ala	Asp	Glu	Leu 85	Phe	Thr	Leu	Val	Leu 90	Pro	Ile	Asn	Ile	Ala 95	Asp	
30	Phe	. Leu	Leu	Arg 100	Gln	Trp	Pro	Phe	Gly 105	Glu	Leu	Met	Cys	Lys 110	Leu	Ile	
	Va]	. Ala	Ile 115	Asp	Gln	Tyr	Asn	Thr 120	Phe	Ser	Ser	Leu	Tyr 125	Phe	Leu	Thr	
	۷a]	. Met	Ser	Ala	Asp	Arg	Tyr	Leu	Val	Val	Leu	Ala	Thr	Ala	Glu	Ser	

		, 130	•				135					140					
	Arg • 145	Arg	Val	Ala	Gly	Arg 150	Thr	Tyr	Ser	Ala	Ala 155	Arg	Ala	Val	Ser	Leu 160	
5	Ala	Val	Trp	Gly	Ile 165	Val	Thr	Leu	Val	Val 170	Leu	Pro	Phe	Ala	Val 175	Phe	
	Ala	Arg	Leu	Asp 180	Asp	Glu	Gln	Gly	Arg 185	Arg	Gln	Cys	Val	Leu 190	Val	Phe	
	Pro	Gln	Pro 195	Glu	Ala	Phe	Trp	Trp 200	Arg	Ala	Ser	Arg	Leu 205	Tyr	Thr	Leu	
10	Val	Leu 210	Gly	Phe	Ala	Ile	Pro 215	Val	Ser	Thr	Ile	Cys 220	Val	Leu	Tyr	Thr	
	Thr 225	Leu	Leu	Cys	Arg	Leu 230	His	Ala	Met	Arg	Leu 235			His		Lys 240	
15	Ala	Leu	Glu	Arg	Ala 245	Lys	Lys	Arg	Val	Lys 250	Phe	Leu	Val	Val	Ala 255	Ile	
	Leu	Ala	Val	Cys 260	Leu	Leu	Cys	Trp	Thr 265	Pro	Tyr	His	Leu	Ser 270	Thr	Val	
	Val	Ala	Leu 275	Thr	Thr	Asp	Leu	Pro 280	Gln	Thr	Pro	Leu	Val 285	Ile	Ala	Ile	
20	Ser	Tyr 290	Phe	Ile	Thr	Ser	Leu 295	Thr	Tyr	Ala	Asn	Ser 300	Cys	Leu	Asn	Pro	:
	Phe	Leu	Туr	Ala	Phe	Leu 310	Asp	Ala	Ser		Arg 315	Arg	Asn	Leu	Arg	Gln 320	
25	Leu	Ile	Thr	Cys	Arg 325	Ala	Ala	Ala				• .	. •				
	(172) IN	FORMA	MOITA	1 FOR	SEÇ	OI O	NO:1	71:									
30	(i)	(B) (C)	LEN TYP STR	CE CH IGTH: PE: n RANDE POLOG	100 ucle DNES	2 ba ic a SS: s	se p cid ingl	airs									
	(i:	i) MC	DLECU	JLE T	YPE:	DNA	(ge	nomi	c)								
	(x:	i) SE	QUEN	ICE D	ESCR	IPTI	ON:	SEQ	ID N	0:17	1:						
	ATGCAGGC	CG CI	GGGC	ACCC	AGA	.GCCC	CTT	GACA	GCAG	ĠG G	CTCC	TTCT	c cc	TCCC	CACG	6	5 C
35	ATGGGTGC	CA AC	GTCT	CTCA	GGA	CAAT	GGC	ACTG	GCCA	CA A	TGCC	ACCT	т ст	CCGA	GCCA	12	2 0
	CTGCCGTTC	CC TC	TATG	TGCT	CCT	GCCC	GCC	GTGT	ACTC	CG G	GATC	TGTG	C TG	TGGG	GCTG	18	3 0

	ACTGGCAACA CGGCCGTCAT CCTTGTAATC CTAAGGGCGC CCAAGATGAA GACGGTGACC 24	0
	AACGTGTTCA TCCTGAACCT GGCCGTCGCC GACGGGCTCT TCACGCTGGT ACTGCCTGTC 300	c
	AACATCGCGG AGCACCTGCT GCAGTACTGG CCCTTCGGGG AGCTGCTCTG CAAGCTGGTG 363	כ
	CTGGCCGTCG ACCACTACAA CATCTTCTCC AGCATCTACT TCCTAGCCGT GATGAGCGTG 420)
5	GACCGATACC TGGTGGTGCT GGCCACCGTG AGGTCCCGCC ACATGCCCTG GCGCACCTAC 480)
	CGGGGGGCGA AGGTCGCCAG CCTGTGTGTC TGGCTGGGCG TCACGGTCCT GGTTCTGCCC 540)
	TTCTTCTCTT TCGCTGGCGT CTACAGCAAC GAGCTGCAGG TCCCAAGCTG TGGGCTGAGC 600)
	TTCCCGTGGC CCGAGCAGGT CTGGTTCAAG GCCAGCCGTG TCTACACGTT GGTCCTGGGC 660	j
	TTCGTGCTGC CCGTGTGCAC CATCTGTGTG CTCTACACAG ACCTCCTGCG CAGGCTGCGG 720	,
10	GCCGTGCGGC TCCGCTCTGG AGCCAAGGCT CTAGGCAAGG CCAGGCGGAA GGTGAAAGTC 780	
	CTGGTCCTCG TCGTGCTGGC CGTGTGCCTC CTCTGCTGGA CGCCCTTCCA CCTGGCCTCT 840	
	GTCGTGGCCC TGACCACGGA CCTGCCCCAG ACCCCACTGG TCATCAGTAT GTCCTACGTC ,900	
	ATCACCAGCC TCACGTACGC CAACTCGTGC CTGAACCCCT TCCTCTACGC CTTTCTAGAT 960	
	GACAACTTCC GGAAGAACTT CCGCAGCATA TTGCGGTGCT GA	
15	(173) INFORMATION FOR SEQ ID NO:172:	
20	(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 333 amino acids(B) TYPE: amino acid(C) STRANDEDNESS:(D) TOPOLOGY: not relevant	
	(ii) MOLECULE TYPE: protein	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:172:	
	Met Gln Ala Ala Gly His Pro Glu Pro Leu Asp Ser Arg Gly Ser Phe 1 5 10 15	
25	Ser Leu Pro Thr Met Gly Ala Asn Val Ser Gln Asp Asn Gly Thr Gly 20 25 30	
	His Asn Ala Thr Phe Ser Glu Pro Leu Pro Phe Leu Tyr Val Leu Leu 35 40 45	
30	Pro Ala Val Tyr Ser Gly Ile Cys Ala Val Gly Leu Thr Gly Asn Thr 50 55 60	
	Ala Val Ile Leu Val Ile Leu Arg Ala Pro Lys Met Lys Thr Val Thr 65 70 75 80	

	I	Asn	Val	Phe	Ile	Leu 85	Asn	Leu	Ala	Val	Ala 90	Asp	Gly	Leu	Phe	Thr 95	Leu
	1	/al	Leu	Pro	Val 100	Asn	Ile	Ala	Glu	His 105	Leu	Leu	Gln	Tyr	Trp 110	Pro	Phe
5	C	3ly	Glu	Leu 115	Leu	Cys	Lys	Leu	Val 120	Leu	Ala	Val	Asp	His 125	Tyr	Asn	Ile
	I	?he	Ser 130	Ser	Ile	Tyr	Phe	Leu 135	Ala	Val	Met	Ser	Val 140	Asp	Arg	Tyr	Leu
10		/al	Val	Leu	Ala	Thr	Val 150	Arg	Ser	Arg	His	Met 155	Pro	Trp	Arg	Thr	Tyr 160
	1	Arg	Gly	Ala	Lys	Val 165	Ala	Ser	Leu	Cys	Val 170	Trp	Leu	Gly	Val	Thr 175	Val
	1	Leu	Val	Leu	Pro 180	Phe	Phe	Ser	Phe	Ala 185	Gly	Val	Tyr	Ser	Asn 190	Glu	Leu
15	C	3ln	Val	Pro 195	Ser	Cys	Gly	Leu	Ser 200	Phe	Pro	Trp	Pro	Glu 205	Gln	Val	Trp
		Phe	Lys 210	Ala	Ser	Arg	Val	Tyr 215.	Thr	Leu	Val	Leu	Gly 220	Phe	Val	Leu	Pro
20		/al 225	Cys	Thr	Ile	Cys	Val 230	Leu	Tyr	Thr	Asp	Leu 235	Leu	Arg	Árg:	Leu	Arg 240
	2	Ala	Val	Arg	Leu	Arg 245	Ser	Gly	Ala	Lys	Ala 250	Leu	Gly	Lys	Ala	Arg 255	Arg
	1	Lys	Val	Lys	Val 260	Leu	Val	Leu	Val	Val 265	Leu	Ala	Val	Cys	Leu 270	·Leu	Cys
25	•	rp	Thr	Pro 275	Phe	His	Leu	Ala	Ser 280	Val	Val	Ala	Leu	Thr 285	Thr	Asp	Leu
	. 1	Pro	Gln 290	Thr	Pro	Leu	Val	Ile 295				Tyr		Ile	Thr	Ser	Leu
30		Thr 305	Tyr	Ala	Asn	Ser	Cys 310	Leu	Asn		Phe	Leu 315	Tyr	Ala	Phe	Leu	Asp 320
	į	Asp	Asn	Phe	Arg	Lys 325	Asn	Phe	Arg	Ser	Ile 330	Leu	Arg	Cys			
	(174)	IN	FORM	ATIO	N FOI	R SE(Q ID	NO:	173:								

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1107 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii)	MOLECULE	TYPE:	DNA	(genomic)
------	----------	-------	-----	-----------

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:173:

	ATGGTCCTTG	AGGTGAGTGA	CCACCAAGTG	CTAAATGACG	CCGAGGTTGC	CGCCCTCCTG	60
	GAGAACTTCA	GCTCTTCCTA	TGACTATGGA	GAAAACGAGA	GTGACTCGTG	CTGTACCTCC	120
5	CCGCCCTGCC	CACAGGACTT	CAGCCTGAAC	TTCGACCGGG	CCTTCCTGCC	AGCCCTCTAC	180
	AGCCTCCTCT	TTCTGCTGGG	GCTGCTGGGC	AACGGCGCGG	TGGCAGCCGT	GCTGCTGAGC	240
	CGGCGGACAG	CCCTGAGCAG	CACCGACACC	TTCCTGCTCC	ACCTAGCTGT	AGCAGACACG	300
	CTGCTGGTGC	TGACACTGCC	GCTCTGGGCA	GTGGACGCTG	CCGTCCAGTG	GGTCTTTGGC	360
	TCTGGCCTCT	GCAAAGTGGC	AGGTGCCCTC	TTCAACATCA	ACTTCTACGC	AGGAGCCCTC	420
10	CTGCTGGCCT	GCATCAGCTT	TGACCGCTAC	CTGAACATAG	TTCATGCCAC	CCAGCTCTAC	480
	CGCCGGGGGC	CCCCGGCCCG	CGTGACCCTC	ACCTGCCTGG	CTGTCTGGGG	GCTCTGCCTG	540
	CTTTTCGCCC	TCCCAGACTT	CATCTTCCTG	TCGGCCCACC	ACGACGAGCG	CCTCAACGCC	600
	ACCCACTGCC	AATACAACTT	CCCACAGGTG	GGCCGCACGG	CTCTGCGGGT	GCTGCAGCTG	660
	GTGGCTGGCT	TTCTGCTGCC	CCTGCTGGTC	ATGGCCTACT	GCTATGCCCA	CATCCTGGCC	720
15	GTGCTGCTGG	TTTCCAGGGG	CCAGCGGCGC	CTGCGGGCCA	AGCGGCTGGT	GGTGGTGGTC	780
	GTGGTGGCCT	TTGCCCTCTG	CTGGACCCCC	TATCACCTGG	TGGTGCTGGT	GGACATCCTC	840
	ATGGACCTGG	GCGCTTTGGC	CCGCAACTGT	GGCCGAGAAA	GCAGGGTAGA	CGTGGCCAAG	900
	TCGGTCACCT	CAGGCCTGGG	CTACATGCAC	TGCTGCCTCA	ACCCGCTGCT	CTATGCCTTT	960
	GTAGGGGTCA	AGTTCCGGGA	GCGGATGTGG	ATGCTGCTCT	TGCGCCTGGG	CTGCCCCAAC	1020
20	CAGAGAGGGC	TCCAGAGGCA	GCCATCGTCT	TCCCGCCGGG	ATTCATCCTG	GTCTGAGACC	1080
	TCAGAGGCCT	CCTACTCGGG	CTTGTGA				1107

(175) INFORMATION FOR SEQ ID NO:174:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 368 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:

- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:174:

	Met 1	Val	Leu	Glu	Val 5	Ser	Asp	His	Gln	Val 10	Leu	Asn	Asp	Ala	Glu 15	Val
• •	Ala	Ala	Leu	Leu 20	Glu	Asn	Phe	Ser	Ser 25	Ser	Tyr	Asp	Tyr	Gly 30	Glu	Asn
5	Glu	Ser	Asp 35	Ser	Cys	Cys	Thr	Ser 40	Pro	Pro	Cys	Pro	Gln 45	Asp	Phe	Ser
	Leu	Asn 50	Phe	Asp	Arg	Ala	Phe 55	Leu	Pro	Ala	Leu	Tyr 60	Ser	Leu	Leu	Phe
10	Leu 65	Leu	Gly	Leu	Leu	Gly 70	Asn	Gly	Ala	Val	Ala 75	Ala	Val	Leu	Leu	Ser 80
•	Arg	Arg	Thr	Ala	Leu 85	Ser	Ser	Thr	Asp	Thr 90	Phe	Leu	Leu	His	Leu 95	Ala
	Val	Ala	Asp	Thr 100	Leu	Leu	Val	Leu	Thr 105	Leu	Pro	Leu,	Trp	Ala 110	Val	Asp
15	Ala	Ala	Val	Gln	Trp	Val	Phe	Gly 120	Ser	Gly	Leu	Cys	Lys 125	Val	Ala	Gly
	Ala	Leu 130	Phe	Asn	Ile	Asn	Phe 135	Tyr	Ala	Gly	Ala	Leu 140	Leu	Leu	Ala	Cys
20	Ile 145	Ser	Phe	Asp	Arg	Tyr 150	Leu	Asn	Ile	Val	His 155	Ala	Thr	Gln	Leu	T yr
	Arg	Arg	Gly	Pro	Pro 165	Ala	Arg	Val	Thr	Leu 170	Thr	Cys	Leu	Ala	Val 175	Trp
	Gly	Leu	Cys	Leu 180	Leu	Phe	Ala	Leu	Pro 185	Asp	Phe	Ile	Phe	Leu 190	Ser	Ala
25	His	His	Asp 195	Glu	Arg	Leu	Asn	Ala 200	Thr	His	Cys	Gln	Tyr 205	Asn	Phe	Pro
	Gln	Val 210	Gly	Arg	Thr	Ala	Leu 215	Arg	Val	Leu	Gln	Leu 220	Val	Ala	Gly	Phe
30	Leu 225	Leu	Pro	Leu	Leu	Val 230	Met	Ala	Tyr	Cys	Tyr 235	Ala	His	Ile	Leu	Ala 240
	Val	Leu	Leu	Val	Ser 245	Arg	Gly	Gln	Arg	Arg 250	Leu	Arg	Ala	Lys	Arg 255	Lev
	Val	Val	Val	Val 260	Val	Val	Ala	Phe	Ala 265	Leu	Cys	Trp	Thr	Pro 270	Tyr	His
35	Leu	Val	Val 275	Leu	Val	Asp	Ile	Leu 280	Met	'Asp	Leu	Gly	Ala 285	Leu	Ala	Arç
	Asn	Cve	Glv	Ara	Glu	Ser	Ara	Va]	Asp	Va)	Ala	Lvs	Ser	Val	Thr	Ser

WO 00/22129 PCT/US99/23938

	2	290		295		• .	300		
	Gly I 305	Leu Gly T		His Cys 310	Cys Leu	Asn Pro 315	Leu Leu	Tyr Ala	Phe 320
5	Val 0	Gly Val L	ys Phe <i>1</i> 325 ·		Arg Met	Trp Met	Leu Leu	Leu Arg	Leu
	Gly C	Cys Pro A	sn Gln <i>1</i> 40		Leu Gln 345	Arg Gln	Pro Ser	Ser Ser 350	Arg
	Arg A	Asp Ser S 355	er Trp S	Ser Glu	Thr Ser	Glu Ala	Ser Tyr 365	Ser Gly	Leu
10	(176) INFC	RMATION	FOR SEQ	ID NO:1	175:				
15	•	SEQUENCE (A) LENG (B) TYPE (C) STRAI (D) TOPO	TH: 1074 : nuclei NDEDNESS LOGY: li	base p.c acid s: singl	airs .e	- .		٠.	÷
		MOLECUL							,
	(xi)	SEQUENC	E DESCRI	PTION:	SEQ ID N	0:175:			
	ATGGCTGATG	ACTATGG	CTC TGAA	TCCACA	TCTTCCAT	GG AAGAC	TACGT TA	ACTTCAAC	· · 60
	TTCACTGACT	TCTACTG	IGA GAAA	AACAAT	GTCAGGCA	GT TTGCG	AGCCA TI	TCCTCCCA	120
20	CCCTTGTACT	GGCTCGT	GTT CATC	GTGGGT	GCCTTGGG	CA ACAGT	CTTGT TA	TCCTTGTC	180
	TACTGGTACT	GCACAAG	AGT GAAG	ACCATG	ACCGACAT	GT TCCTT	TTGAA TT	TGGCAATT	240
	GCTGACCTCC	TCTTTCT	TGT CACT	CTTCCC	TTCTGGGC	CA TTGCT	GCTGC TG	ACCAGTGG	300
	AAGTTCCAGA	CCTTCATO	GTG CAAG	GTGGTC	AACAGCAT	GT ACAAG	ATGAA CT	TCTACAGC	360
	TGTGTGTTGC	TGATCATO	GTG CATC	AGCGTG	GACAGGTA	CA TTGCC	ATTGC CC	AGGCCATG	420
25	AGAGCACATA	CTTGGAG	GGA GAAA	AGGCTT	TTGTACAG	CA AAATG	GTTTG CT	TTACCATC	480
	TGGGTATTGG	CAGCTGCT	CT CTGC	ATCCCA	GAAATCTT.	AT ACAGC	CAAAT CA	AGGAGGAA	540
	TCCGGCATTG	CTATCTG	CAC CATG	GTTTAC	CCTAGCGA'	TG AGAGC	ACCAA AC	TGAAGTCA	600
	GCTGTCTTGA	CCCTGAAC	GT CATT	CTGGGG	TTCTTCCT	TC CCTTC	GTGGT CA	TGGCTTGC	660
	TGCTATACCA	TCATCATT	CA CACC	CTGATA	CAAGCCAA	GA AGTCT	TCCAA GC	ACAAAGCC	720
30	AAGAAAGTGA	CCATCACT	GT CCTG	ACCGTC	TTTGTCTT	GT CTCAG	TTTCC CT	ACAACTGC	780
	ATTTTGTTGG	TGCAGACO	CAT TGAC	GCCTAT	GCCATGTT	CA TCTCC	AACTG TG	CCGTTTCC	840

	ACCAACATT	G ACATC	TGCTT C	CAGGTCAC	CC CAG	ACCATCG	CCTTCTTC	CA CAGTI	GCCTG	900
	AACCCTGTT	C TCTAT	GTTTT TO	etgggtg <i>i</i>	AG AGA	TTCCGCC	GGGATCTC	GT GAAAA	ACCCTG	960
	AAGAACTTG	G GTTGC	ATCAG C	TAGGCCCA	AG TGG	GTTTCAT	TTACAAGG	AG AGAGG	GAAGC	1020
	TTGAAGCTG'	T CGTCT	ATGTT G	CTGGAGAC	CA ACC	TCAGGAG	CACTCTCC	CT CTGA		1074
5	(177) INF	ORMATIO	N FOR SI	EQ ID NO):176:					
10	(i)	(A) LEI (B) TYI (C) STI	CE CHARI NGTH: 39 PE: amin RANDEDNI POLOGY:	7 amino no acid ESS:	acid					
	(ii)) MOLECI	ULE TYPE	: prote	in					
	(xi)) SEQUEI	NCE DESC	RIPTION	I: SEQ	ID. NO:	L76:		. •	
	Met 1	Ala Asp	Asp Tyr	Gly Se	er Glu	Ser Thi	Ser Ser	Met Glu	Asp 1	Tyr
15	Val <i>I</i>	Asn Phe	Asn Phe	Thr As	p Phe	Tyr Cys	Glu Lys	Asn Asn 30		Arg
	Gln I	Phe Ala 35	Ser His	Phe Le	u Pro 40	Pro Le	Tyr Trp	Leu Val	Phe	Ile
20		Gly Ala 50	Leu Gly	Asn Se		Val Ile	Leu Val	Tyr Trp	Tyr	Cys
	Thr <i>1</i> 65	Arg Val	Lys Thr	Met Th		Met Phe	Leu Leu 75	Asn Leu	Ala	
	Ala A	Asp Leu	Leu Phe	Leu Va	l Thr	Leu Pro	Phe Trp	Ala Ile	Ala i 95	Ala
25	Ala A	Asp Gln	Trp Lys	Phe Gl	n Thr	Phe Met	Cys Lys	Val Val 110	Asn :	Ser
	Met 1	Tyr Lys 115	Met Asn	Phe Ty	r Ser 120	Cys Val	Leu Leu	Ile Met 125	Cys :	Ile
30		Val Asp	Arg Tyr	Ile Al 13		Ala Glr	Ala Met 140	Arg Ala	His !	Th <u>r</u>
	Trp <i>I</i> 145	Arg Glu	Lys Arg	Leu Le 150	u Tyr	Ser Lys	Met Val	Cys Phe		Ile 160
	Trp \	/al Leu	Ala Ala 165		u Cys	Ile Pro	Glu Ile	Leu Tyr	Ser (Gln
35	Ile I	Lys Glu	Glu Ser 180	Gly Il	e Ala	Ile Cys	Thr Met	Val Tyr 190	Pro S	Ser

,

٠	Asp	Glu	Ser 195		Lys	Leu	Lys	Ser 200	Ala	Val	Leu	Thr	Leu 205	Lys	Val	Ile
	Lev	Gly 210		Phe	Leu	Pro	Phe 215	Val	Val	Met	Ala	Cys 220	Cys	Tyr	Thr	Ile
5	Ile 225	Ile	His	Thr	Leu	Ile 230	Gln	Ala	Lys	Lys	Ser 235	Ser	Lys	His	Lys	Ala 240
	Lys	Lys	Val	Thr	Ile 245	Thr	Val	Leu	Thr	Val 250	Phe	Val	Leu	Ser	Gln 255	Phe
10	Pro	Tyr	Asn	Cys 260	Ile	Leu	Leu	Val	Gln 265	Thr	Ile	Asp	Ala	туr 270	Ala	Met
	Phe	Ile	Ser 275	Asn	Cys	Ala	Val	Ser 280	Thr	Asn	Ile	Asp	Ile 285	Cys	Phe	Gln
	Val	Thr 290	Gln	Thr	Ile	Ala	Phe 295	Phe	His	Ser	Cys	Leu 300	Asn	Pro	Val	Leu
15	Tyr 305	Val	Phe	Val	Gly	Glu 310	Arg	Phe	Arg	Arg	Asp 315	Leu	Val	Lys	Thr	Leu 320
	Lys	Asn	Leu	Gly	Cys 325	Ile	Ser	Gln	Ala	Gln 330	Trp	Val	Ser	Phe	Thr 335	Arg
20	Arg	Glu	Gly	Ser 340	Leu	Lys	Leu	Ser	Ser 345	Met	Leu	Leu	Glu	Thr 350	Thr	Ser
	Gly	Ala	Leu 355	Ser	Leu				•							
	(178) IN	FORM	AOITA	1 FOR	SEQ	D	NO:1	.77:								
25	(i	(B)	LEN TYP STR	CE CH IGTH: PE: n RANDE POLOG	111 ucle DNES	0 ba ic a S: s	se p cid ingl	airs								
	(i:	i) MC	LECU	LE T	YPE:	DNA	. (ge	nomi	c)							
30	(x:	i) SE	QUEN	ICE D	ESCR	IPTI	ON:	SEQ	ID N	0:17	7:					
	ATGGCCTC	AT CO	ACCA	CTCG	GGG	CCCC	AGG	GTTT	CTGA	CT T	ATTT	TCTG	G GC	TGCC	GCCG	60
	GCGGTCAC	AA CI	CCCG	CCAA	CCA	GAGC	GCA	GAGG	CCTC	GG C	GGGC	AACG	G GT	CGGT	GGCT	120
	GGCGCGGA	CG CI	'CCAG	CCGT	CAC	GCCC	TTC	CAGA	GCCT	GC A	GCTG	GTGC	A TC	AGCT	GAAG	180
	GGGCTGATO															
35	CTGGTGCT	G TG	ATCG	CGCG	GGT	GCCG	CGG	CTGC.	ACAA	CG T	GACG.	AACT'	r cc	TCAT	CGGC	300

	AACCTGGCCT TGTCCGACGT GCTCATGTGC ACCGCCTGCG TGCCGCTCAC GCTGGCCTAT	360
	GCCTTCGAGC CACGCGGCTG GGTGTTCGGC GGCGGCCTGT GCCACCTGGT CTTCTTCCTG	420
	CAGCCGGTCA CCGTCTATGT GTCGGTGTTC ACGCTCACCA CCATCGCAGT GGACCGCTAC	480
	GTCGTGCTGG TGCACCCGCT GAGGCGCGCA TCTCGCTGCG CCTCAGCCTA CGCTGTGCTG	540
5	GCCATCTGGG CGCTGTCCGC GGTGCTGGCG CTGCCGCCCG CCGTGCACAC CTATCACGTG	600
	GAGCTCAAGC CGCACGACGT GCGCCTCTGC GAGGAGTTCT GGGGCTCCCA GGAGCGCCAG	660
	CGCCAGCTCT ACGCCTGGGG GCTGCTGCTG GTCACCTACC TGCTCCCTCT GCTGGTCATC	720
	CTCCTGTCTT ACGTCCGGGT GTCAGTGAAG CTCCGCAACC GCGTGGTGCC GGGCTGCGTG	780
	ACCCAGAGCC AGGCCGACTG GGACCGCGCT CGGCGCCGGC GCACCAAATG CTTGCTGGTG	840
10	GTGGTCGTGG TGGTGTTCGC CGTCTGCTGG CTGCCGCTGC ACGTCTTCAA CCTGCTGCGG	900
	GACCTCGACC CCCACGCCAT CGACCCTTAC GCCTTTGGGC TGGTGCAGCT GCTCTGCCAC	960
	TGGCTCGCCA TGAGTTCGGC CTGCTACAAC CCCTTCATCT ACGCCTGGCT GCACGACAGC	1020
	TTCCGCGAGG AGCTGCGCAA ACTGTTGGTC GCTTGGCCCC GCAAGATAGC CCCCCATGGC	1080
	CAGAATATGA CCGTCAGCGT GGTCATCTGA	1110
15	(179) INFORMATION FOR SEQ ID NO:178:	٠.
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 369 amino acids (B) TYPE: amino acid	
20	(C) STRANDEDNESS: (D) TOPOLOGY: not relevant	
	(ii) MOLECULE TYPE: protein.	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:178:	
	Met Ala Ser Ser Thr Thr Arg Gly Pro Arg Val Ser Asp Leu Phe Se	r
25	Gly Leu Pro Pro Ala Val Thr Thr Pro Ala Asn Gln Ser Ala Glu Al 20 25 30	a .
	Ser Ala Gly Asn Gly Ser Val Ala Gly Ala Asp Ala Pro Ala Val Th	r

Pro Phe Gln Ser Leu Gln Leu Val His Gln Leu Lys Gly Leu Ile Val

Leu Leu Tyr Ser Val Val Val Val Gly Leu Val Gly Asn Cys Leu

	Le:	u Vai	l Le	u Val	1 Ile 85	e Ala	Arg	y Val	Pro	90	j Lei	His	s Ası	n Va	1 Th: 95	r Asn
	Phe	e Lei	ı Ile	e Gly 100	/ Asr	ı Leu	Ala	. Lev	Ser 105	Asp	Val	Leu	ı Met	Cy:		Ala
5	Cys	s Val	115	Leu 5	ı Thr	Leu	Ala	120	Ala	Phe	e Glu	Pro) Arc		y Trp	Val
	Phe	130	/ Gly	/ Gly	Leu .	Cys	His 135	Leu	Val	Phe	Phe	Leu 140		Pro	Val	Thr
10	Val	Tyr	Val	. Ser	Val	Phe 150	Thr	Leu	Thr	Thr	Ile 155	Ala	Val	Asp	Arg	Tyr 160
	Val	Val	Leu	Val	His 165	Pro	Leu	Arg	Arg	Ala 170		Arg	Cys	Ala	Ser 175	Ala
	Tyr	Ala	Val	Leu 180	Ala	Ile	Trp	Ala	Leu 185	Ser	Ala	Val	Leu	Ala 190		Pro
15	Pro	Ala	Val 195	His	Thr	Tyr	His	Val 200	Glu	Leu	Lys	Pro	His 205	Asp	Val	Arg
	Leu	Cys 210	Glu	Glu	Phe	Trp	Gly 215	Ser	Gln	Glu	Arg	Gln 220	Arg	Gln	Leu	Tyr
20	Ala 225	Trp	Gly	Leu	Leu	Leu 230	Val	Thr	Tyr	Leu	Leu 235	Pro	Leu	Leu	Val	Ile 240
	Leu	Leu	Ser	Tyr	Val 245	Arg	Val	Ser	Val	Lys 250		Arg	Asn	Arg	Val 255	Val "
	Pro	Gly	Cys	Val 260	Thr	Gln	Ser	Gln	Ala 265	Asp	Trp	Asp	Arg	Ala 270	Arg	Arg
25	Arg	Arg	Thr 275	Lys	Cys	Leu	Leu	Val 280	Val	Val	Val	Val	Val 285	Phe	Ala	Val
	Cys	Trp 290	Leu	Pro	Leu	His	Val 295	Phe	Asn	Leu	Leu	Arg 300	Asp	Leu	Asp :	Pro
30	His 305	Ala	Ile	Asp	Pro	Tyr 310	Ala	Phe	Gly		Val 315	Gln	Leu	Leu	Cys	His 320
	Trp	Leu	Ala	Met	Ser 325	Ser	Ala	Cys		Asn 330	Pro	Phe	Ile	Tyr	Ala 335	Trp
	Leu	His	Asp	Ser 340	Phe	Arg	Glu		Leu 345	Arg	Lys .	Leu		Val 350	Ala	Trp
35	Pro	Arg	Lys 355	Ile .	Ala	Proj		Gly (360	Gln .	Asn i	Met '		Val 365	Ser	Val	Val
	-1.															

(180) INFORMATION FOR SEQ ID NO:179:

5

(i)	SEQUENCE	CHARACTERISTICS:
-----	----------	------------------

- (A) LENGTH: 1083 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:179:

10	ATGGACCCAG	AAGAAACTTC	AGTTTATTTG	GATTATTACT	ATGCTACGAG	CCCAAACTCT	60
	GACATCAGGG	AGACCCACTC	CCATGTTCCT	TACACCTCTG	TCTTCCTTCC	AGTCTTTTAC	120
	ACAGCTGTGT	TCCTGACTGG	AGTGCTGGGG	AACCTTGTTC	TCATGGGAGC	GTTGCATTTC	180
	AAACCCGGCA	GCCGAAGACT	GATCGACATC	TTTATCATCA	ATCTGGCTGC	CTCTGACTTC	240
	ATTTTTCTTG	TCACATTGCC	TCTCTGGGTG	GATAAAGAAG	CATCTCTAGG	ACTGTGGAGG	300
15	ACGGGCTCCT	TCCTGTGCAA	AGGGAGCTCC	TACATGATCT	CCGTCAATAT	GCACTGCAGT	360
	GTCCTCCTGC	TCACTTGCAT	GAGTGTTGAC	CGCTACCTGG	CCATTGTGTG	GCCAGTCGTA	420
	TCCAGGAAAT	TCAGAAGGAC	AGACTGTGCA	TATGTAGTCT	GTGCCAGCAT	CTGGTTTATC	480
	TCCTGCCTGC	TGGGGTTGCC	TACTCTTCTG	TCCAGGGAGC	TCACGCTGAT	TGATGATAAG	540
	CCATACTGTG	CAGAGAAAAA	GGCAACTCCA	ATTAAACTCA	TATGGTCCCT	GGTGGCCTTA	600
20	ATTTTCACCT	TTTTTGTCCC	TTTGTTGAGC	ATTGTGACCT	GCTACTGTTG	CATTGCAAGG	660
	AAGCTGTGTG	CCCATTACCA	GCAATCAGGA	AAGCACAACA	AAAAGCTGAA	GAAATCTAAG	720
	AAGATCATCT	TTATTGTCGT	GGCAGCCTTT	CTTGTCTCCT	GGCTGCCCTT	CAATACTTTC	780
	AAGTTCCTGG	CCATTGTCTC	TGGGTTGCGG	CAAGAACACT	ATTTACCCTC	AGCTATTCTT	840
	CAGCTTGGTA	TGGAGGTGAG	TGGACCCTTG	GCATTTGCCA	ACAGCTGTGT	CAACCCTTTC	900
25	ATTTACTATA	TCTTCGACAG	CTACATCCGC	CGGGCCATTG	TCCACTGCTT	GTGCCCTTGC	960.
	CTGAAAAACT	ATGACTTTGG	GAGTAGCACT	GAGACATCAG	ATAGTCACCT	CACTAAGGCT	1020
	CTCTCCACCT	TCATTCATGC	AGAAGATTTT	GCCAGGAGGA	GGAAGAGGTC	TGTGTCACTC	1080
	TAA					-	1083

_ ____

(181) INFORMATION FOR SEQ ID NO:180:

30 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 360 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein

	(i	i) M	OLEC	ULE	TYPE	: pr	otei	.n								
5	(x	i) S	EQUE	NCE	DESC	RIPT	'ION :	SEÇ	ID	NO : 1	.80:					
`	Met 1	Asp	Pro	Glu	Glu 5	Thr	Ser	Val	Tyr	Leu 10		Туг	Tyr	Tyr	Ala 15	Thr
	Ser	Pro	Asn	Ser 20	Asp	Ile	Arg	Glu	Thr 25	His	Ser	His	Val	Pro 30	Tyr	Thr
10	Ser	Val	Phe 35	Leu	Pro	Val	Phe	Туr 40	Thr	Ala	Val	Phe	Leu 45	Thr	Gly	Val
	Leu	Gly 50	Asn	Leu	Val	Leu	Met 55	Gly	Ala	Leu	His	Phe 60	Lys	Pro	Gly	Ser
15	Arg 65	Arg	Leu	Ile	Asp	Ile 70	Phe	Ile	Ile	Asn	Leu 75	Ala	Ala	Ser	Asp	Phe 80
·	Ile	Phe	Leu	Val	Thr 85	Leu	Pro	Leu	Trp	Val 90	Asp	Lys	Glu	Ala	Ser 95	Leu:
	Gly	Leu	Trp	Arg 100	Thr	Gly	Ser	Phe	Leu 105	Cys	Lys	Gly	Ser	Ser 110	Tyr	Met
20	Ile	Ser	Val 115	Asn	Met	His	Cys	Ser 120	Val	Leu	Leu	Leu	Thr 125	Cys	Met	Ser
		Asp 130	Arg	Tyr	Leu	Ala	Ile 135	Val	Trp	Pro	Val	Val 140	Ser	Arg	Lys	Phe
25	Arg 145	Arg	Thr	Asp	Cys	Ala 150	Tyr	Val	Val	Cys	Ala 155	Ser	Ile	Trp	Phe	Ile 160
	Ser	Cys	Leu	Leu	Gly 165	Leu	Pro	Thr	Leu	Leu 170	Ser	Arg	Glu	Leu	Thr 175	Leu
20	Ile			180					185					190		
30	٠		195					200			Thr		205			
		210					215				Ala	220				
35	His 225					230			•	**	235					240
	Lys	Ile	Ile	Phe	Ile 245	Val	Val	Ala	Ala	Phe 250	Leu	Val	Ser	Trp	Leu 255	Pro

	Phe	Asn	Thr	Phe 260	Lys	Phe	Leu	Ala	Ile 265	Val	Ser	Gly	Leu	Arg 270	Gln	Glu	
	His	Tyr	Leu 275	Pro	Ser	Ala	Ile	Leu 280	Gln	Leu	Gly	Met	Glu 285	Val	Ser	Gly	
5	Pro	Leu 290	Ala	Phe	Ala	Asn	Ser 295	Cys	.Val	Asn	Pro	Phe 300	Ile	Tyr	Tyr	Ile	
	Phe 305	Asp	Ser	Tyr	Ile	Arg 310	Arg	Ala	Ile	Val	His 315	Суѕ	Leu	Cys	Pro	Cys 320	
10	Leu	Lys	Asn	Tyr	Asp 325	Phe	Gly	Ser	Ser	Thr 330	Glu	Thr	Ser	Asp	Ser 335	His	
	Leu	Thr	Lys	Ala 340	Leu	Ser	Thr	Phe	Ile 345	His	Ala	Glu	Asp	Phe 350	Ala	Arg	
	Arg	Arg	Lys 355	Arg	Ser	Val	Ser	Leu 360			•		٠.				
15	(182) IN	FORM	OITA	N FOI	R SE	Q ID	NO:	181:					٠				
20	(i.	(B)	LEI TYI TYI TOI	NGTH PE: 1 RANDI POLOG	: 10: nucle EDNE: GY: :	20 basic a SS: : linea : DN	ase] acid sing: ar A (g	pair: le enom	ic)	NO:1	81:		· .				
	ATGAATGG	CÇ T	TGAA	GTGG	C TC	CCCC.	AGGT	CTG	ATCA	CCA ·	ACTT	CTCC	CT G	GCCA	CGGC	A.	60
	GAGCAATG'	TG G	CCAG	GAGA(C GC	CACT	GGAG	AAC.	ATGC'	TGT	TCGC	CTCC'	TT C			G	120
25	GATTTTAT	CC T	GGCT'	TTAGʻ	: T TG	GCAA	TACC	CTG	GCTC	TGT	GGCT	TTTC.	AT C	CGAG		С	180
	AAGTCCGG	GA C	CCCG	GCCA	A CG	TGTT	CCTG	ATG	CATC	TGG	CCGT	GGCC	GA C	TTGT	CGTG	C	240
	GTGCTGGT	CC T	GCCC.	ACCC(G CC	TGGT	CTAC	CAC	TTCT	CTG	GGAA	CCAC'	TG G	CCAT	TTGG	G	300
	GAAATCGC	AT G	CCGT	CTCA	C CG	GCTT	CCTC	TTC	TACC	TCA	ACAT	GTAC	GC C	AGCA'	TCTA	С	360
	TTCCTCAC	CT G	CATC	AGCG	C CG	ACCG	TTTC	CTG	GCCA	TTG	TGCA	CCCG	GT C	AAGT	CCCT	С	420
30	AAGCTCCG	CA G	GCCC	CTCT.	A CG	CACA	.CCTG	GCC	TGTG	CCT	TCCT	GTGG	GT G	GTGG	TGGC	Т	480
	GTGGCCAT	GG C	CCCG	CTGC	T GG	TGAG	CCCA	CAG	ACCG	TGC	AGAC	CAAC	CA C	ACGG	TGGT	С	540
	TGCCTGCA	GC T	GTAC	CGGG	A GA	AGGC	CTCC	CAC	CATG	CCC	TGGT	GTCC	CT G	GCAG	TGGC	С	600
	TTCACCTT	cc c	GTTC	ATCA	C CA	.CGGT	CACC	TGC	TACC	TGC	TGAT	CATC	CG C	AGCC	TGCG	G	660

WO 00/22129 PCT/US99/23938

133

CAGGGCCTGC	GTGTGGAGAA	GCGCCTCAAG	ACCAAGGCAA	AACGCATGAT	CGCCATAGTG	720
CTGGCCATCT	TCCTGGTCTG	CTTCGTGCCC	TACCACGTCA	ACCGCTCCGT	CTACGTGCTG	780
CACTACCGCA	GCCATGGGGC	CTCCTGCGCC	ACCCAGCGCA	TCCTGGCCCT	GGCAAACCGC	840
ATCACCTCCT	GCCTCACCAG	CCTCAACGGG	GCACTCGACC	CCATCATGTA	TTTCTTCGTG	900
GCTGAGAAGT	TCCGCCACGC	CCTGTGCAAC	TTGCTCTGTG	GCAAAAGGCT	CAAGGGCCCG	960
CCCCCAGCT	TCGAAGGGAA	AACCAACGAG	AGCTCGCTGA	GTGCCAAGTC	AGAGCTGTGA	1020
·(183) INFOR	RMATION FOR	SEQ ID NO:	L82:			

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 339 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:

5

10

- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:182: 15 Met Asn Gly Leu Glu Val Ala Pro Pro Gly Leu Ile Thr Asn Phe Ser 10 Leu Ala Thr Ala Glu Gln Cys Gly Gln Glu Thr Pro Leu Glu Asn Met Leu Phe Ala Ser Phe Tyr Leu Leu Asp Phe Ile Leu Ala Leu Val Gly 20 Asn Thr Leu Ala Leu Trp Leu Phe Ile Arg Asp His Lys Ser Gly Thr Pro Ala Asn Val Phe Leu Met His Leu Ala Val Ala Asp Leu Ser Cys 75 25 Val Leu Val Leu Pro Thr Arg Leu Val Tyr His Phe Ser Gly Asn His 90 Trp Pro Phe Gly Glu Ile Ala Cys Arg Leu Thr Gly Phe Leu Phe Tyr 105 Leu Asn Met Tyr Ala Ser Ile Tyr Phe Leu Thr Cys Ile Ser Ala Asp 30 120
 - Pro Leu Tyr Ala His Leu Ala Cys Ala Phe Leu Trp Val Val Ala
 145 150 155 160

Arg Phe Leu Ala Ile Val His Pro Val Lys Ser Leu Lys Leu Arg Arg

Val Ala Met Ala Pro Leu Leu Val Ser Pro Gln Thr Val Gln Thr Asn

					100	• • • •	TÓD		• • •	5 .		170	٠.		• • • •	. ,	175		
			His	Thr	Val	Val 180	Cys	Leu	Gln	Leu	Tyr 185	Arg	Glu	Lys	Ala	Ser 190	His	His	
	5		Ala	Leu	Val 195		Leu	Ala	Val		Phe			Pro	Phe 205	Ile	Thr	Thr	
	-		Val	Thr 210	Cys	Tyr	Leu	Leu	Ile 215	Ile	Arg	Ser	Leu	Arg 220	Gln	Gly	Leu	Arg	
			Val 225	Glu	Lys	Arg	Leu	Lys 230	Thr	Lys	Ala		Arg 235	Met	Ile	Ala	Ile	Val 240	
	10		Leu	Ala	Ile	Phe	Leu 245	Val	Cys	Phe	Val	Pro 250	Tyr	His	Val	Asn	Arg 255	Ser	
		•	Val	Tyr							His 265	_			_			Gln	
	15		Arg	Ile	Leu 275	Ala	Leu	Ala	Asn	Arg 280	Ile	Thr.	Ser	Cys	Leu 285	Thr	Ser	Leu	
		, ;	Asn	Gly 290	Ala	Leu	Asp	Pro	Ile 295	Met	Tyr	Phe	Phe	Val 300	Ala	Glu	Lys	Phe	
•	. , ,		Arg 305	His	Ala	Leu	Cys	Asn 310	Leu	Leu	Cys	Gly	Lys 315	Arg	Leu	Lys	Gly	Pro 320	
;	20	* v	Pro	Pro	Ser	Phe	Glu 325	Gly	Lys	Thr	Asn	Glu 330	Ser	Ser	Leu	Ser	Ala 335	Lys	
			Ser	Glu	Leu						•	:				.•			
	25	(183)		FORMA				•					. •				. •		
·			. (- /	(A) (B) (C)	LEN TYP STR	GTH: PE: 1 PEANDE	996 nucle EDNES	bas ic a S: s	se pa cid singl	airs					ţ		•		
•	30		(ii	i) MC	DLECU	ne i	YPE:	DNA	A (ge	enomi	.c)								
				i) SE															
		TACA																	
		ATCA																	
	35	ATGAT	rgaat	rg To	GCAT	TAGI	GGA	CTTG	SATA	TTTA	TAAT	GA C	TTTA	CCCI	T TC	GAAT	'GTTT		

	TATTATGCAA AAGATGAATG GCCATTTGGA GAGTACTTCT GCCAGATTCT TGGAGCTCTC	300
	ACAGTGTTTT ACCCAAGCAT TGCTTTATGG CTTCTTGCCT TTATTAGTGC TGACAGATAC	360
	ATGGCCATTG TACAGCCGAA GTACGCCAAA GAACTTAAAA ACACGTGCAA AGCCGTGCTG	420
	GCGTGTGTGG GAGTCTGGAT AATGACCCTG ACCACGACCA CCCCTCTGCT ACTGCTCTAT	480
5	AAAGACCCAG ATAAAGACTC CACTCCCGCC ACCTGCCTCA AGATTTCTGA CATCATCTAT	540
	CTAAAAGCTG TGAACGTGCT GAACCTCACT CGACTGACAT TTTTTTTCTT GATTCCTTTG	600
	TTCATCATGA TTGGGTGCTA CTTGGTCATT ATTCATAATC TCCTTCACGG CAGGACGTCT	660
	AAGCTGAAAC CCAAAGTCAA GGAGAAGTCC AAAAGGATCA TCATCACGCT GCTGGTGCAG	720
	GTGCTCGTCT GCTTTATGCC CTTCCACATC TGTTTCGCTT TCCTGATGCT GGGAACGGGG	780
10	GAGAATAGTT ACAATCCCTG GGGAGCCTTT ACCACCTTCC TCATGAACCT CAGCACGTGT	840
	CTGGATGTGA TTCTCTACTA CATCGTTTCA AAACAATTTC AGGCTCGAGT CATTAGTGTC	900
	ATGCTATACC GTAATTACCT TCGAAGCATG CGCAGAAAAA GTTTCCGATC TGGTAGTCTA	960
	AGGTCACTAA GCAATATAAA CAGTGAAATG TTATGA	996
	(185) INFORMATION FOR SEQ ID NO:184:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 331 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: (D) TOPOLOGY: not relevant 	
20	(ii) MOLECULE TYPE: protein	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:184:	
ž	Met Ile Thr Leu Asn Asn Gln Asp Gln Pro Val Pro Phe Asn Ser Ser 1 5 10 15	
25	His Pro Asp Glu Tyr Lys Ile Ala Ala Leu Val Phe Tyr Ser Cys Ile 20 25 30	
	Phe Ile Ile Gly Leu Phe Val Asn Ile Thr Ala Leu Trp Val Phe Ser 35 40 45	
	Cys Thr Thr Lys Lys Arg Thr Thr Val Thr Ile Tyr Met Met Asn Val 50 55 60	
30	Ala Leu Val Asp Leu Ile Phe Ile Met Thr Leu Pro Phe Arg Met Phe 65 70 75 80	
	Tyr Tyr Ala Lys Asp Glu Trp Pro Phe Gly Glu Tyr Phe Cys Gln Ile 85 90 95	

		Leu	Gly	Ala	Leu 100	Thr	Val	Phe	Tyr	Pro 105	Ser	Ile	Ala	Leu	Trp. 110	Leu	Leu
		Ala	Phe	Ile 115	Ser	Ala	Asp	Arg	Tyr 120	Met	Ala	Ile	Val	Gln 125	Pro	Lys	Tyr
5		Ala	Lys 130	Glu	Leu	Lys	Asn	Thr 135	Cys	Lys	Ala	Val	Leu 140	Ala	Cys	Val	Gly
		Val 145	_	Ile	Met	Thr	Leu 150	Thr	Thr	Thr	Thr	Pro 155	Leu	Leu	Leu	Leu	Tyr 160
10		Lys	Asp	Pro	Asp	Lys 165	Asp	Ser	Thr	Pro	Ala 170	Thr	Cys	Leu	Lys	Ile 175	Ser
		Asp	Ile	Ile	Туг 180	Leu	Lys	Ala	Val	Asn 185	Val	Leu	Asn	Leu	Thr 190	Arg	Leu
÷		Thr	Phe	Phe 195	Phe	Leu	Ile	Pro	Leu 200	Phe	Ile	Met	Ile	Gly 205	Cys	Tyr	Leu
15			Ile 210	Ile	His	Asn	Leu	Leu 215	His	Gly	Arg	Thr	Ser 220	Lys	Leu	Lys	Pro
		Lys 225	Val	Lys	Glu	Lys	Ser 230	Lys	Arg	Ile	Ile	Ile 235	Thr	Leu	Leu	Val	Gln 240
20		Val	Leu	Val	Cys	Phe 245	Met	Pro	Phe	His	Ile 250	Cys	Phe	Ala	Phe	Leu 255	Met
		Leu	Gly	Thr	Gly 260		Asn	Ser	Tyr	Asn 265		Trp	Gly	Ala	Phe 270	Thr	Thr
		Phe	Leu	Met 275		Leu	Ser	Thr	Cys 280		Asp	Val		Leu 285		Tyr	Ile
25		Val	Ser 290		Gln	Phe	Gln	Ala 295		Val	Ile	Ser	Val 300		. Leu	.Tyr	Arg
			Tyr		Arg	Ser		Arg			Ser	Phe 315		Ser	Gly	Ser	Leu 320
30		Arg	Ser	Leu	Ser	Asn 325	Ile	Asn	Ser	Glu	Met 330						
	(186) IN					Q ID										
35		(1	(F (E	A) LE B) TY C) SI	NGTH PE: RANI	i: 10 nucl	77 b leic ESS: line	ase ació sing	pair l	s							

(ii) MOLECULE TYPE: DNA (genomic)

	(xi)	SEQUENCE D	ESCRIPTION:	SEQ ID NO:	185:	to the second	
	ATGCCCTCTG	TGTCTCCAGC	GGGGCCCTCG	GCCGGGGCAG	TCCCCAATGC	CACCGCAGTG	6
	ACAACAGTGC	GGACCAATGC	CAGCGGGCTG	GAGGTGCCCC	TGTTCCACCT	GTTTGCCCGG	12
	CTGGACGAGG	AGCTGCATGG	CACCTTCCCA	GGCCTGTGCG	TGGCGCTGAT	GGCGGTGCAC	180
5	GGAGCCATCT	TCCTGGCAGG	GCTGGTGCTC	AACGGGCTGG	CGCTGTACGT	CTTCTGCTGC	240
	CGCACCCGGG	CCAAGACACC	CTCAGTCATC	TACACCATCA	ACCTGGTGGT	GACCGATCTA	300
	CTGGTAGGGC	TGTCCCTGCC	CACGCGCTTC	GCTGTGTACT	ACGGCGCCAG	GGGCTGCCTG	360
	CGCTGTGCCT	TCCCGCACGT	CCTCGGTTAC	TTCCTCAACA	TGCACTGCTC	CATCCTCTTC	420
	CTCACCTGCA	TCTGCGTGGA	CCGCTACCTG	GCCATCGTGC	GGCCCGAAGG	CTCCCGCCGC	480
10	TGCCGCCAGC	CTGCCTGTGC	CAGGGCCGTG	TGCGCCTTCG	TGTGGCTGGC	CGCCGGTGCC	540
	GTCACCCTGT	CGGTGCTGGG	CGTGACAGGC	AGCCGGCCCT	GCTGCCGTGT	CTTTGCGCTG	600
	ACTGTCCTGG	AGTTCCTGCT	GCCCCTGCTG	GTCATCAGCG	TGTTTACCGG	CCGCATCATG	660
	TGTGCACTGT	CGCGGCCGGG	TCTGCTCCAC	CAGGGTCGCC	AGCGCCGCGT	GCGGGCCAAG	720
	CAGCTCCTGC	TCACGGTGCT	CATCATCTTT	CTCGTCTGCT	TCACGCCCTT	CCACGCCCGC	7,80
15	CAAGTGGCCG	TGGCGCTGTG	GCCCGACATG	CCACACCACA	CGAGCCTCGT	GGTCTACCAC	840
	GTGGCCGTGA	CCCTCAGCAG	CCTCAACAGC	TGCATGGACC	CCATCGTCTA	CTGCTTCGTC	900
	ACCAGTGGCT	TCCAGGCCAC	CGTCCGAGGC	CTCTTCGGCC	AGCACGGAGA	GCGTGAGCCC	960
	AGCAGCGGTG 2	ACGTGGTCAG	CATGCACAGG	AGCTCCAAGG	GCTCAGGCCG	TCATCACATC	1020
	CTCAGTGCCG-0	GCCCTCACGC	CCTCACCCAG	GCCCTGGCTA	ATGGGCCCGA	GGCTTAG	1077
20	(187) INFOR	MATION FOR	SEQ ID NO:1	86:			
			RACTERISTIC 358 amino a			•	
	(E	3) TYPE: am:	ino acid	Club			

- (C) STRANDEDNESS:
- 25 (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:186:

Met Pro Ser Val Ser Pro Ala Gly Pro Ser Ala Gly Ala Val Pro Asn 10 15

30 Ala Thr Ala Val Thr Thr Val Arg Thr Asn Ala Ser Gly Leu Glu Val 25

	Pro	Leu	Phe 35	His	Leu	Phe	Ala	Arg 40	Leu	Asp	Glu	Glu	Leu 45	His	Gly	Thr
	Phe	Pro 50	Gly	Leu	Cys	Val	Ala 55	Leu	Met	Ala	Val	His 60	Gly	Ala	Ile	Phe
5	Leu 65	Ala	Gly	Leu	Val	Leu 70	Asn	Gly	Leu	Ala	Leu 75	Tyr	Val	Phe	Cys	Cys 80
	Arg	Thr	Arg	Ala	Lys 85	Thr	Pro	Ser	Val	Ile 90	Tyr	Thr	Ile	Asn	Leu 95	Val
10	Val	Thr	Asp	Leu 100	Leu	Val	Gly	Leu	Ser 105	Leu	Pro	Thr	Arg	Phe 110	Ala	Val
	Tyr	Tyr	Gly 115	Ala	Arg	Gly	Cys	Leu 120	Arg	Cys	Ala	Phe	Pro 125	His	Val	Leu
	Gly	Tyr 130	Phe	Leu	Asn	Met	His 135	Cys	Ser	Ile	Leu	Phe 140	Leu-	Thr	Cys	Ile
15	Cys 145	Val	Asp	Arg	Tyr	Leu 150	Ala	Ile	Val	Arg	Pro 155	Glu	Gly	Ser	Arg	Ala 160
	Cys	Arg	Gln	Pro	Ala 165	Cys	Ala	Arg	Ala	Val 170	Cys	Ala	Phe	Val	Trp 175	Leu
20	Ala	Ala	Gly	Ala 180	Val	Thr	Leu	Ser	Val 185	Leu	Gly	Val	Thr	Gly 190		Arg
	Pro	Cys	Cys 195	Arg	Val	Phe	Ala	Leu 200	Thr	Val	Leu	Glu	Phe 205	Leu	Leu	Pro
	Leu	Leu 210	Val	Ile	Ser	Val	Phe 215	Thr	Gly	Arg	Ile	Met 220	Cys	Ala	Leu	Ser
25	Arg 225	Pro	Gly	Leu	Leu	His 230	Gln	Gly	Arg	Gln	Arg 235	Arg	Val	Arg	Ala	Lys 240
	Gln	Leu	Leu	Leu	Thr 245	Val	Leu	Ile	Ile	Phe 250	Leu	Val	Cys	Phe	Thr 255	Pro
30	Phe	His	Ala	Arg 260	Gln	Val	Ala	Val	Ala 265	Leu	Trp	Pro	Asp	Met 270	Pro	His
	His	Thr	Ser 275	Leu	Val	Val	Tyr	His 280		Ala	Val	Thr	Leu 285		Ser	Leu
		290	Cys				295					300				
35	Gln 305	Ala	Thr	Val	Arg	Gly 310	Leu	Phe			315					320
	_	_	~ 3	•	77-7	77-7	0	31-4	TT4 -	7	C	C ~ ~	Taro	~7.	Cor	<i>a</i> 1,

10

Arg His His Ile Leu Ser Ala Gly Pro His Ala Leu Thr Gln Ala Leu 340 345

Ala Asn Gly Pro Glu Ala

(188) INFORMATION FOR SEQ ID NO:187:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1050 base pairs
 (B) TYPE: nucleic acid

 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:187:

	ATGAACTCCA	CCTTGGATGG	TAATCAGAGC	AGCCACCCTT	TTTGCCTCTT	GGCATTTGGC	, 6 ·
15	TATTTGGAAA	CTGTCAATTT	TTGCCTTTTG	GAAGTATTGA	TTATTGTCTT	TCTAACTGTA	
	TTGATTATTT	CTGGCAACAT	CATTGTGATT	TTTGTATTTC	ACTGTGCACC	TTTGTTGAAC	18
-	CATCACACTA	CAAGTTATTT	TATCCAGACT	ATGGCATATG	CTGACCTTTT	TGTTGGGGTG	240
	AGCTGCGTGG	TCCCTTCTTT	ATCACTCCTC	CATCACCCC	TTCCAGTAGA	GGAGTCCTTG	300
	ACTTGCCAGA	TATTTGGTTT	TGTAGTATCA	GTTCTGAAGA	GCGTCTCCAT	GGCTTCTCTG	360
20	GCCTGTATCA	GCATTGATAG	ATACATTGCC	ATTACTAAAC	CTTTAACCTA	TAATACTCTG	420
	GTTACACCCT	GGAGACTACG	CCTGTGTATT	TTCCTGATTT	GGCTATACTC	GACCCTGGTC	480
	TTCCTGCCTT	CCTTTTTCCA	CTGGGGCAAA	CCTGGATATC	ATGGAGATGT	GTTTCAGTGG	540
	TGTGCGGAGT	CCTGGCACAC	CGACTCCTAC	TTCACCCTGT	TCATCGTGAT	GATGTTATAT	600
	GCCCCAGCAG	CCCTTATTGT	CTGCTTCACC	TATTTCAACA	TCTTCCGCAT	CTGCCAACAG	660
25	CACACAAAGG	ATATCAGCGA	AAGGCAAGCC	CGCTTCAGCA	GCCAGAGTGG	GGAGACTGGG	720
	GAAGTGCAGG	CCTGTCCTGA	TAAGCGCTAT	AAAATGGTCC	TGTTTCGAAT	CACTAGTGTA	780
	TTTTACATCC	TCTGGTTGCC	ATATATCATC	TACTTCTTGT	TGGAAAGCTC	CACTGGCCAC	840
	AGCAACCGCT	TCGCATCCTT	CTTGACCACC	TGGCTTGCTA	TTAGTAACAG	TTTCTGCAAC	900
	TGTGTAATTT	ATAGTCTCTC	CAACAGTGTA	TTCCAAAGAG	GACTAAAGCG	CCTCTCAGGG	960
30	GCTATGTGTA	CTTCTTGTGC	AAGTCAGACT	ACAGCCAACG	ACCCTTACAC	AGTTAGAAGC	1020
	AAAGGCCCTC	TTAATGGATG	TCATATCTGA				1050

And the control of th (189) INFORMATION FOR SEQ ID NO:188: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 349 amino acids (B) TYPE: amino acid 5 (C) STRANDEDNESS: (D) TOPOLOGY: not relevant (ii) MOLECULE TYPE: protein (xi) SEQUENCE DESCRIPTION: SEQ ID NO:188: Met Asn Ser Thr Leu Asp Gly Asn Gln Ser Ser His Pro Phe Cys Leu 10 10 Leu Ala Phe Gly Tyr Leu Glu Thr Val Asn Phe Cys Leu Leu Glu Val 25 Leu Ile Ile Val Phe Leu Thr Val Leu Ile Ile Ser Gly Asn Ile Ile 15 Val Ile Phe Val Phe His Cys Ala Pro Leu Leu Asn His His Thr Thr 50 55 Ser Tyr Phe Ile Gln Thr Met Ala Tyr Ala Asp Leu Phe Val Gly Val 70 75 80 Ser Cys Val Val Pro Ser Leu Ser Leu Leu His His Pro Leu Pro Val 20 90 Glu Glu Ser Leu Thr Cys Gln Ile Phe Gly Phe Val Val Ser Val Leu 100 105 Lys Ser Val Ser Met Ala Ser Leu Ala Cys Ile Ser Ile Asp Arg Tyr 115 120 25 Ile Ala Ile Thr Lys Pro Leu Thr Tyr Asn Thr Leu Val Thr Pro Trp 135 Arg Leu Arg Leu Cys Ile Phe Leu Ile Trp Leu Tyr Ser Thr Leu Val 150 155 . . . Phe Leu Pro Ser Phe Phe His Trp Gly Lys Pro Gly Tyr His Gly Asp 30 165 170 Val Phe Gln Trp Cys Ala Glu Ser Trp His Thr Asp Ser Tyr Phe Thr 185

Phe Thr Tyr Phe Asn Ile Phe Arg Ile Cys Gln Gln His Thr Lys Asp 210 215 220

Leu Phe Ile Val Met Met Leu Tyr Ala Pro Ala Ala Leu Ile Val Cys 195 200 205

	Ile 225	Ser	Glu	Arg	Gln	Ala 230	Arg	Phe	Ser	Ser	Gln 235	Ser	Gly	Glu	Thr	Gly 240	
	Glu	Val	Gln	Ala	Cys 245	Pro	Asp	Lys	Arg	Tyr 250	Lys	Met	Val	Leu	Phe 255	Arg	
5	Ile	Thr	Ser	Val 260	Phe	Tyr	Ile	Leu	Trp 265	Leu	Pro	туr	Ile	Ile 270	Tyr	Phe	
	Leu	Leu	Glu 275	Ser	Ser	Thr	Gly	His 280	Ser	Asn	Arg	Phe	Ala 285	Ser	Phe	Leu	
10	Thr	Thr 290	Trp	Leu	Ala	Ile	Ser 295	Asn	Ser	Phe	Cys	Asn 300	Cys	Val	Ile	Tyr	
	Ser 305	Leu	Ser	Asn	Ser	Val 310	Phe	Gln	Arg		Leu 315	Lys	Arg	Leu	Ser	Gly 320	
	Ala	Met	Cys	Thr	Ser 325	Cys	Ala	Ser	Gln	Thr 330	Thr	Ala	Asn	Asp	Pro 335	Tyr	
15	Thr	Val	Arg	Ser 340	Lys	Gly	Pro	Leu	Asn 345	Gly	Cys	His	Ile		· ,	. 7	
	(190) IN	FORM	ATION	N FOR	SEC) ID	NO:1	89:						•			
20	(i)	(A) (B) (C)	LEN TYI STF	CE CH NGTH: PE: T RANDE	130 nucle	02 ba eic a SS: s	se p cid singl	airs	3			٠.				:	
	(ii			POLOG				nomi	.c)								
	. (x i	i) SE	EQUEN	NCE I	ESCR	RIPTI	ON:	SEQ	ID N	IO:18	9:		4	•	,		
25	ATGTGTTT	rt Ci	CCCA	ATTCI	GGA	AATC	CAAC	ATGC	AGTO	TG A	ATCI	'AACA	TA	CAGI	'GCGA	. 6	•
	GATGACATT	rg at	GAC#	ATCAA	CAC	CAAI	ATG	TACC	AACC	AC I	'ATCA	TATO	C GI	TAAG	CTTI	12	(
	CAAGTGTCT	rc To	CACCG	GATT	TCI	TATG	ATT	GAAA	TTGT	GT I	'GGGA	CTTG	G CA	GCAA	CCTC	18	•
	ACTGTATTO	G TA	CTTI	CACTG	CAT	GAAA	TCC	AACI	'TAAT	CA A	CTCT	'GTCA	G TA	ACAI	TATI	24	(
	ACAATGAAT	rc ti	CATO	TACT	TGA	TGTA	ATA	ATTI	GTGT	'GG G	ATGT	'ATTC	C TC	TAAC	TATA	30	(
30	GTTATCCTT	C TG	CTTI	CACI	GGA	GAGT	'AAC	ACTG	CTCT	'CA I	TTGC	TGTT	T CC	ATGA	.GGCT	36	(
	TGTGTATCT	TT TI	GCAA	GTGT	CTC	AACA	GCA	ATCA	ACGT	TT T	TGCT	'ATCA	C TI	'TGGA	CAGA	42	(
	TATGACATO	T CI	GTAA	AACC	TGC	AAAC	CGA.	ATTC	TGAC	'AA I	GGGC	AGAG	C TG	TAAT	GTTA	48	(
	ATGATATCO	TT A	TGGA	\TTTT	TTC	TTTT	TTC	TCTT	TCCT	GA T	TCCT	TTTA	T TG	AGGT	AAAT	54	(
	TTTTTCAGI	C TI	CAAA	GTGG	AAA	TACC	TGG	GAAA	ACAA	GA C	ACTT	TTAT	G TG	TCAG	TACA	60	(

	AATGAATA	ACT A	CACI	GAAC	T GO	GAAI	GTAT	' TAI	CACC	TGT	TAG	ACAG	AT (CCCAA	TATT	'C	660
	TTTTTCAC	CTG I	TGTA	GTAA	T GI	TAAT	CACA	TAC	ACCA	AAA	TACT	TCAG	GC 1	rctt <i>r</i>	ATAT	T	720
	CGAATAGO	GCA C	AAGA	TTTT	C AA	CAGG	GCAG	AAG	AAGA	AAG	CAAG	AAAG	AA A	AAAGA	CAAT	T	780
	TCTCTAAC	CA C	ACAA	CATG.	A GG	CTAC	AGAC	ATG	TCAC	AAA	GCAG	TGGT	'GG C	SAGAA	ATGT	Α	840
5	GTCTTTGG	STG I	'AAGA	ACTT	C AG	TTTC	TGTA	ATA	ATTG	CCC	TCCG	GCGA	GC 1	rgtga	AACG	A	900
	CACCGTGA	AC G	ACGA	GAAA	G AC	AAAA	GAGA	GTC	AAGA	.GGA	TGTC	TTTA	TT G	ATTA	TTTC	T	960
	ACATTTCT	TC T	CTGC	TGGA	CAC	CAAT	TTCT	GTT	TTAA	ATA	CCAC	CATT	TT A	TGTT	TAGG	C 1	020
	CCAAGTGA	CC T	TTTA	GTAA	TA A	TAAG	ATTG	TGT	TTTT	TAG	TCAT	GGCT	TA I	GGAA	CAAC'	Т 1	080
•	ATATTTCA	rcc c	TCTA	TTAT	A TG	CATT	CACT	AGA	CAAA	TAA	TTCA	AAAG	GT C	TTGA	AAAG'	T 1:	140
10	AAAATGAA	AA A	.GCGA	GTTG:	r TT	CTAT	AGTA	GAA	GCTG	ATC	CCCT	GCCT	AA T	'AATG	CTGT	A 1	200
÷	ATACACAA	CT C	TTGG	ATAG	A TC	CCAA	AAGA	AAC	АААА	AAA	TTAC	CTTT	GA A	GATA	GTGA	A 1:	260
	ATAAGAGA	AA A	ACGT	TTAG:	r GC	CTCA	GGTT	GTC	ACAG	ACT	AG					13	302
	(191) IN	FORM	ATIO:	N FOI	SE	Q ID	NO:	190:									
15	(i			CE CE						• :		-		•			
				PE: a			id										
				POLO			rele	vant				٠.					
	(i	i) M	OLEC	ULE 1	YPE	: pr	otei	n									
20	(x	i) S	EQUE	NCE I	ESC:	RIPT	ION:	SEQ	ID I	NO:1	90:						
	Met 1	Cys	Phe	Ser	Pro 5	Ile	Leu	Glu	Ile	Asn 10	Met	Gln	Ser	Glu	Ser 15	Asn	•
	Ile	Thr	Val	Arg 20	Asp	Asp	Ile	Asp	Asp 25	Ile	Asn	Thr	Asn	Met 30	Tyr	Gln	
25	Pro	Leu	Ser 35	Tyr	Pro	Leu	Ser	Phe 40	Gln	Val	Ser	Leu	45	Gly		Leu	
	Met	Leu 50	Glu	Ile	Val	Leu	Gly 55	Leu	Gly	Ser	Asn	Leu 60		Val		Val	-
30	Leu 65	Tyr	Cys	Met	Lys	Ser 70	Asn	Leu	Ile	Asn	Ser 75	Val	Ser	Asn	Ile	Ile 80	
	Thr	Met	Asn	Leu	His 85	Val	Leu	Asp	Val	Ile 90	Ile	Cys	Val	Gly	Cys 95	Ile	

Pro Leu Thr Ile Val Ile Leu Leu Ser Leu Glu Ser Asn. Thr Ala

materiorini viain monatana i las

	•		٠.		100	• ′	<i>*</i>	~ ·		105				-	110		
		Leu	Ile	Cys 115	Cys	Phe	His	Glu	Ala 120	Cys	Val	Ser	Phe	Ala 125	Ser	Val	Ser
5		Thr	Ala 130		Asn		Phe			Thr	Leu	Asp	Arg 140	Tyr	Asp	Ile	Ser
		Val 145	Lys	Pro	Ala	Asn	Arg 150	Ile	Leu	Thr	Met	Gly 155	Arg	Ala	Val	Met	Leu 160
		Met	Ile	Ser	Ile	Trp 165	Ile	Phe	Ser	Phe	Phe 170		Phe	Leu	Ile	Pro 175	Phe
10		Ile	Glu	Val	Asn 180	Phe	Phe	Ser	Leu	Gln 185	Ser	Gly	Asn	Thr	Trp 190	Glu	Asn
		Lys		Leu 195			Val			Asn					Glu		Gly
15		Met	Tyr 210	Tyr	His	Leu	Leu	Val 215	Gln	Ile	Pro	Ile	Phe 220	Phe	Phe	Thr	Va'l
		Val 225	Val	Met	Leu	Ile	Thr 230	Tyr	Thr	Lys	Ile	Leu 235	Gln	Ala	Leu	Asn	Ile 240
		Arg '	Ile	Gly	Thr	Arg 245	Phe	Ser	Thr	Gly	Gln 250	Lys	Lys	Lys	Ala	Arg 255	Lys
20		Lys	Lys	Thr	Ile 260	Ser	Leu	Thr	Thr	Gln 265	His		Ala	Thr	Asp 270	Met	Ser
		Gln	Ser	Ser 275	Gly	Gly	Arg	Asn	Val 280	Val	Phe	Gly	Val	Arg 285	Thr	Ser	Val
25								Arg 295							Arg		Arg
		Arg 305	Glu	Arg	Gln	Lys	Arg 310	Val	Lys	Arg	Met	Ser 315	Leu	Leu	Ile	Ile	Ser 320
			Phe		Leu			Thr			Ser 330	Val	Leu	Asn	Thr		Ile
30		Leu	Cys	Leu	Gly 340	Pro	Ser	Asp	Leu	Leu 345	Val	Lys	Leu	Arg	Leu 350	Cys	Phe
		Leu		Met 355	Ala	Tyr	Gly	Thr	Thr 360	Ile	Phe	His	Pro		Leu	_	Ala
35		Phe	Thr 370	Arg	Gln	Lys	Phe	Gln 375	Lys	Val	Leu	Lys	Ser 380	Lys	Met	Lys	Lys
		Arg 385	Val		Ser		Val 390	Glu	Ala	Asp		Leu 395		Asn		Ala	Val 400

Ile His Asn Ser Trp Ile Asp Pro Lys Arg Asn Lys Lys Ile Thr Phe
405 410 415

Glu Asp Ser Glu Ile Arg Glu Lys Arg Leu Val Pro Gln Val Val Thr 420 425 430

5 Asp

(192) INFORMATION FOR SEQ ID NO:191:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1209 base pairs
- 10 (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:191:

ATGTTGTGTC CTTCCAAGAC AGATGGCTCA GGGCACTCTG GTAGGATTCA CCAGGAAACT 60 CATGGAGAAG GGAAAAGGGA CAAGATTAGC AACAGTGAAG GGAGGGAGAA TGGTGGGAGA 120 GGATTCCAGA TGAACGGTGG GTCGCTGGAG GCTGAGCATG CCAGCAGGAT GTCAGTTCTC 180 AGAGCAAAGC CCATGTCAAA CAGCCAACGC TTGCTCCTTC TGTCCCCAGG ATCACCTCCT 240 CGCACGGGGA GCATCTCCTA CATCAACATC ATCATGCCTT CGGTGTTCGG CACCATCTGC 300 20 CTCCTGGGCA TCATCGGGAA CTCCACGGTC ATCTTCGCGG TCGTGAAGAA GTCCAAGCTG 360 CACTGGTGCA ACAACGTCCC CGACATCTTC ATCATCAACC TCTCGGTAGT AGATCTCCTC TTTCTCCTGG GCATGCCCTT CATGATCCAC CAGCTCATGG GCAATGGGGT GTGGCACTTT GGGGAGACCA TGTGCACCCT CATCACGGCC ATGGATGCCA ATAGTCAGTT CACCAGCACC TACATCCTGA CCGCCATGGC CATTGACCGC TACCTGGCCA CTGTCCACCC CATCTCTTCC ACGAAGTTCC GGAAGCCCTC TGTGGCCACC CTGGTGATCT GCCTCCTGTG GGCCCTCTCC TTCATCAGCA TCACCCCTGT GTGGCTGTAT GCCAGACTCA TCCCCTTCCC AGGAGGTGCA GTGGGCTGCG GCATACGCCT GCCCAACCCA GACACTGACC TCTACTGGTT CACCCTGTAC 780 CAGTTTTTCC TGGCCTTTGC CCTGCCTTTT GTGGTCATCA CAGCCGCATA CGTGAGGATC CTGCAGCGCA TGACGTCCTC AGTGGCCCCC GCCTCCCAGC GCAGCATCCG GCTGCGGACA AAGAGGGTGA AACGCACAGC CATCGCCATC TGTCTGGTCT TCTTTGTGTG CTGGGCACCC 30 TACTATGTGC TACAGCTGAC CCAGTTGTCC ATCAGCCGCC CGACCCTCAC CTTTGTCTAC TTATACAATG CGGCCATCAG CTTGGGCTAT GCCAACAGCT GCCTCAACCC CTTTGTGTAC 1080

	ATCGTGCT	CT (GTGA	GACG'	TT C	CGCA	AACG	C TT	GGTC	CTGT	CGG	TGAA	GCC	TGĊA	GĆĆC	AG :	1140
	GGGCAGCT	TC (GCGC'	TGTC	AG C	AACG	CTCA	G AC	GGCT	GACG	AGG.	AGAG	GAC	AGAA	AGCA	AA :	1200
	GGCACCTC	A]	L209
•	(193) IN	FORM	ITAL	ON FO	OR SI	EQ II	ON C	:192	:								
5	(i	(<i>I</i> (E	A) LE 3) TY C) ST	engti (PE : [rani	H: 40 amir DEDNE	02 ar 10 ac ESS:	RISTI mino cid rele	acio									
10	(i	i) M	OLEC	TULE	TYPE	: pı	otei	.n									
	(x	i) s	EQUE	NCE	DESC	RIPI	: NOI	SEC) ID	NO : 1	92:				-		
	Met 1	Leu	Cys	Pro	Ser 5	Lys	Thr	Asp	Gly	Ser 10	Gly	/ His	Ser		Arg	, Ile	
15	His	Gln	Glu	Thr 20	His	Gly	Glu	Gly	Lys 25	Arg	Asp	Lys	Ile			Ser	
	Glu	Gly	Arg 35	Glu	Asn	Gly	Gly	Arg 40	Gly	Phe	Gln	Met	Asn 45	Gly	Gly	Ser	i.
	Leu	Glu 50	Ala	Glu	His	Ala	Ser 55	Arg	Met	Ser	Val	Leu 60	Arg	Ala	Lys	Pro	•
20	65					70					75					Pro 80	
					85					90					95	Phe	
25				100					105			Ser		110			
			115					120				Asn	125				
		130					135					Leu 140				_	
30	Met 145					150					155			•		160	
	Gly				165					170					175		
35	Phe '			180					185					190	•		
	Ala '	Thr	Val	His	Pro	Ile	Ser	Ser	Thr	Lys	Phe	Arg	Lvs	Pro	Ser	Val	

	٠,		195					200	٠			<i>:</i>	205				
	Ala	Thr 210	Leu	Val	Ile	Cys	Leu 215	Leu	Trp	Ala	Leu	Ser 220	Phe	Ile	Ser	Ile	
5	Thr 225	Pro	Val	Trp	Leu	Tyr 230	Ala	Arg	Leu	Ile	Pro 235	Phe	Pro	Gly	Gly	Ala 240	
	Val	Gly	Cys	Gly	Ile 245	Arg	Leu	Pro	Asn	Pro 250	Asp	Thr	Asp	Leu	Tyr 255	Trp	
	Phe	Thr	Leu	Tyr 260	Gln	Phe	Phe	Leu	Ala 265	Phe	Ala	Leu	Pro	Phe 270	Val	Val	
10	Ile	Thr	Ala 275	Ala	Tyr	Val	Arg	Ile 280	Leu	Gln	Arg	Met	Thr 285	Ser	Ser	Val	
	Ala	Pro 290	Ala		Gln	Arg	Ser 295	Ile	Arg	Leu	Arg	Thr 300	Lys	Arg	Val	Lys	
15	Arg 305		Ala	Ile	Ala	Ile 310	Cys	Leu	Val	Phe	Phe 315	Val	Cys	Trp	Ala	Pro 320	•
	Tyr	Tyr	Val	Leu	Gln 325	Leu	Thr	Gln	Leu	Ser 330	Ile	Ser	Arg	Pro	Thr 335	Leu	
	Thr	Phe	Val	Tyr 340	Leu	Tyr	Asn	Ala	Ala 345	Ile	Ser	Leu	Gly	Tyr 350	Ala	Asn	٠.
20		-	355					360					365			Arg	
	-	370	1				375					380				Arg	
25	Ala 385		. Ser	Asn	Ala	Gln 390		· Ala	Asp	Glu	Glu 395	Arg	Thr	Glu	Ser	Lys 400	
	Gly	Thr	:														
	(194) IN	1FORI	MTIC	n fo	R SE	Q II	NO:	193:					(#)				
20	(:	i) SI	EQUEN	ICE C	HARA	CTEF	RISTI	CS:									
30		(I	3) TY	PE:	nucl	.eic	acid	1					**				
	1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	· (I	C) ST C) TO	POLC	EDNE GY:	ESS: line	sing ear	Jle 	K w	, .							
	(:	ii) ľ	MOLEC	CULE	TYPE	E: Di	() A	genor	nic)								
35	(:	xi) :	SEQUI	ENCE	DESC	CRIP	NOIT	: SE(Q ID	NO: 1	193:						
	ATGGATG'	TGA (CTTC	CAAC	C C	CGGG	GCGT(G GG	CCTG	SAGA	TGT	ACCCI	AGG (CACCO	GCGCI	łC	60
	GCTGCGG	ccc (CCAA	CACC	AC C	rccc	CCGA	G CT	CAAC	CTGT	CCC	ACCC	GCT (CTG	GCA(C	120

	GCCCTGGCCA ATGGGACAGG TGAGCTCTCG GAGCACCAGC AGTACGTGAT CGGCCTGTTC	180
	CTCTCGTGCC TCTACACCAT CTTCCTCTTC CCCATCGGCT TTGTGGGCAA CATCCTGATC	240
	CTGGTGGTGA ACATCAGCTT CCGCGAGAAG ATGACCATCC CCGACCTGTA CTTCATCAAC	300
	CTGGCGGTGG CGGACCTCAT CCTGGTGGCC GACTCCCTCA TTGAGGTGTT CAACCTGCAC	36.0
5	GAGCGGTACT ACGACATCGC CGTCCTGTGC ACCTTCATGT CGCTCTTCCT GCAGGTCAAC	420
	ATGTACAGCA GCGTCTTCTT CCTCACCTGG ATGAGCTTCG ACCGCTACAT CGCCCTGGCC	480
	AGGGCCATGC GCTGCAGCCT GTTCCGCACC AAGCACCACG CCCGGCTGAG CTGTGGCCTC	540
	ATCTGGATGG CATCCGTGTC AGCCACGCTG GTGCCCTTCA CCGCCGTGCA CCTGCAGCAC	600
	ACCGACGAGG CCTGCTTCTG TTTCGCGGAT GTCCGGGAGG TGCAGTGGCT CGAGGTCACG	660
10	CTGGGCTTCA TCGTGCCCTT CGCCATCATC GGCCTGTGCT ACTCCCTCAT TGTCCGGGTG	720
	CTGGTCAGGG CGCACCGGCA CCGTGGGCTG CGGCCCCGGC GGCAGAAGGC GAAACGCATG	780
	ATCCTCGCGG TGGTGCTGGT CTTCTTCGTC TGCTGGCTG	840
	GTGCACCTCC TGCAGCGGAC GCAGCCTGGG GCCGCTCCCT GCAAGCAGTC TTTCCGCCAT	900
	GCCCACCCC TCACGGGCCA CATTGTCAAC CTCGCCGCCT TCTCCAACAG CTGCCTAAAC	960
15	CCCCTCATCT ACAGCTTTCT CGGGGAGACC TTCAGGGACA AGCTGAGGCT GTACATTGAG	1020
	CAGAAAACAA ATTTGCCGGC CCTGAACCGC TTCTGTCACG CTGCCCTGAA GGCCGTCATT	1080
	CCAGACAGCA CCGAGCAGTC GGATGTGAGG TTCAGCAGTG CCGTGTGA	1128
	(195) INFORMATION FOR SEQ ID NO:194:	
20	(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 375 amino acids(B) TYPE: amino acid(C) STRANDEDNESS:(D) TOPOLOGY: not relevant	-
	(ii) MOLECULE TYPE: protein	
25	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:194:	
	Met Asp Val Thr Ser Gln Ala Arg Gly Val Gly Leu Glu Met Tyr Pro 1 5 10 15	•

Leu Ser His Pro Leu Leu Gly Thr Ala Leu Ala Asn Gly Thr Gly Glu
35 40 45

Gly Thr Ala His Ala Ala Ala Pro Asn Thr Thr Ser Pro Glu Leu Asn

	Leu	Ser 50	Glu	His	Gln (Tyr 55	Val	TTE	GIÀ	Leu	60	neu	361	Cys	Deu .	
	Tyr 65	Thr	Ile	Phe		Phe 70	Pro	Ile	Gly	Phe	Val 75	Gly	Asn	Ile	Leu	Ile 80	
5	Leu	Val	Val	Asn	Ile 85	Ser	Phe	Arg	Glu	Lys 90	Met	Thr	Ile	Pro	Asp 95	Leu	
	Tyr	Phe	Ile	Asn 100	Leu	Ala	Val	Ala	Asp 105	Leu	Ile	Leu	Val	Ala 110	Asp	Ser	
10	Leu	Ile	Glu 115	Val	Phe	Asn	Leu	His 120	Glu	Arg	Tyr	Tyr	Asp 125	Ile	Ala	Val	
	Leu	Cys 130	Thr	Phe	Met	Ser	Leu 135	Phe	Leu	Gln	Val	Asn 140	Met	Tyr	Ser	Ser	
	Val 145		Phe	Leu	Thr	Trp 150	Met	Ser	Phe	Asp	Arg 155	Tyr	Ile	Ala	Leu	Ala 160	
15	Arg	Ala	Met	Arg	Cys 165	Ser	Leu	Phe	Arg	Thr 170	Lys	His	His	Ala	Arg 175	Leu	
	Ser	Cys	Gly	Leu 180	Ile	Trp	Met	Ala	Ser 185	Val	Ser	Ala	Thr	Leu 190	Val	Pro.	
20	Phe	Thr	Ala 195		His	Leu	Gln	His 200	Thr	Asp	Glu	Ala	Cys 205	Phe	Cys	Phe	
	Ala	Asp 210		Arg	Glu	Val	Gln 215	Trp	Leu	Glu	ı Val	Thr 220	Leu	Gly	Phe	Ile	
,	. 225	5				230					s Ser 235	•	• •			. 240	
25	Let	ı Val	Arg	, Ala	His 245	Arg	His	Arg	, Gly	/ Let 250	ı Arg	Pro	Arg	Arg	Gln 255	Lys	
	Ala	a Lys	arg	9 Met 260		Lev	a Ala	Val	. Val		u Val	Phe	Phe	270	. Cys	Trp	
30			275	5				280)				285			Gln	
		29	0				295	5 :				300)		•	Leu	
	30	5				31	0				31!	5				320	
35	Pr	o Le	u Il		325	5 ,				33	0				33:		
	_	_	+3	- 01		. T.32	e ጥካ	r Ac	n T.e	u Pr	o Al	a Le	u Ası	n Ar	g Ph	e Cys	

340 345 350 His Ala Ala Leu Lys Ala Val Ile Pro Asp Ser Thr Glu Gln Ser Asp Val Arg Phe Ser Ser Ala Val 5 375 370 (196) INFORMATION FOR SEQ ID NO:195: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 960 base pairs (B) TYPE: nucleic acid 10 (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (genomic) (xi) SEQUENCE DESCRIPTION: SEQ ID NO:195: ATGCCATTCC CAAACTGCTC AGCCCCCAGC ACTGTGGTGG CCACAGCTGT GGGTGTCTTG 60 CTGGGGCTGG AGTGTGGGCT GGGTCTGCTG GGCAACGCGG TGGCGCTGTG GACCTTCCTG 120 15 TTCCGGGTCA GGGTGTGGAA GCCGTACGCT GTCTACCTGC TCAACCTGGC CCTGGCTGAC 180 CTGCTGTTGG CTGCGTGCCT GCCTTTCCTG GCCGCCTTCT ACCTGAGCCT CCAGGCTTGG 240 CATCTGGGCC GTGTGGGCTG CTGGGCCCTG CGCTTCCTGC TGGACCTCAG CCGCAGCGTG 300 GGGATGGCCT TCCTGGCCGC CGTGGCTTTG GACCGGTACC TCCGTGTGGT CCACCCTCGG 360 CTTAAGGTCA ACCTGCTGTC TCCTCAGGCG GCCCTGGGGG TCTCGGGCCT CGTCTGGCTC 420 CTGATGGTCG CCCTCACCTG CCCGGGCTTG CTCATCTCTG AGGCCGCCCA GAACTCCACC 480 AGGTGCCACA GTTTCTACTC CAGGGCAGAC GGCTCCTTCA GCATCATCTG GCAGGAAGCA 540 600 CTCTCCTGCC TTCAGTTTGT CCTCCCCTTT GGCCTCATCG TGTTCTGCAA TGCAGGCATC ATCAGGGCTC TCCAGAAAAG ACTCCGGGAG CCTGAGAAAC AGCCCAAGCT TCAGCGGGCC 660 720 25 AAGGCACTGG TCACCTTGGT GGTGGTGCTG TTTGCTCTGT GCTTTCTGCC CTGCTTCCTG GCCAGAGTCC TGATGCACAT CTTCCAGAAT CTGGGGAGCT GCAGGGCCCT TTGTGCAGTG 780 GCTCATACCT CGGATGTCAC GGGCAGCCTC ACCTACCTGC ACAGTGTCGT CAACCCCGTG 840 GTATACTGCT TCTCCAGCCC CACCTTCAGG AGCTCCTATC GGAGGGTCTT CCACACCCTC 900 CGAGGCAAAG GGCAGGCAGC AGAGCCCCCA GATTTCAACC CCAGAGACTC CTATTCCTGA 960 30 (197) INFORMATION FOR SEQ ID NO:196:

(i) SEQUENCE CHARACTERISTICS:

A STATE OF THE STA

) TY	PE: RAND	amin EDNE	.9 am lo ac ISS: not	id			t _e .		. •				
5	(i	.i) M	OLEC	ULE	TYPE	: pr	otei	n								
	(x	i) S	EQUE	NCE	DESC	RIPT	ION:	SEÇ	ID	NO:1	96 :	•				
	Met 1	Pro	Phe	Pro	Asn 5	Cys	Ser	Ala	Pro	Ser 10	Thr	Val	Val	Ala	Thr 15	Ala
10	Val	Gly	Val	Leu 20	Leu	Gly	Leu	Glu	Cys 25	Gly	Leu	Gly	Leu	Leu 30	Gly	Asn
	Ala	Val	Ala 35	Leu	Trp	Thr	Phe	Leu 40	Phe	Arg	Val	Arg	Val 45	Trp	Lys	Pro
		50					55					60				Ala
15	65	Cys				70					75					80
		Leu			85					90			, .		95	
20		Arg		100	.,				105					110		
		Leu	115					120					125			
		Ala 130					135			*.		140				
25	145	Thr				150					155					160
		Cys			165					170					175	
30		Gln		180					185					190		
	•	Val	195					200					205			• *
		Glu 210					215					220				
35	Thr 225	Leu	Val	Val	Val	Leu 230	Phe	Ala	Leu	Cys	Phe 235	Leu	Pro	Cys	Phe	Leu 240
	Ala	Arg	Val	Leu	Met	His	Ile	Phe	Gln	Asn	Lev	Glv.	Ser	Cvs	Ara	Δla

			245	: .	- 25	0 ~				255	
	Leu (Cys Ala Val 260	Ala His	Thr Ser	Asp Va 265	l Thr	Gly		Leu 270	Thr	Tyr
5	Leu E	His Ser Val 275	Val Asn	Pro Val		r Cys		Ser 285	Ser	Pro	Thr
		Arg Ser Ser 290	Tyr Arg	Arg Val	Phe Hi	s Thr	Leu :	Arg	Gly	Lys	Gly
	Gln A	Ala Ala Glu	Pro Pro 310	Asp Phe	Asn Pr	o Arg 315	Asp :	Ser '	Tyr	Ser	
10	(198) INFC	RMATION FOR	SEQ ID	NO:197:							
15		SEQUENCE CH (A) LENGTH: (B) TYPE: I (C) STRANDE (D) TOPOLOG MOLECULE T	1143 ba nucleic a DNESS: s Y: linea	se pair cid single cr		·		•		·	÷
	(xi)	SEQUENCE D	ESCRIPTI	ON: SEQ	ID NO:	197:			,		1.5
	ATGGAGGAAG	GTGGTGATTT	TGACAAC	TAC TAT	GGGGCAG	ACAAC	CAGTO	TGA	AGTG'	TGAG	60
	TACACAGACT	GGAAATCCTC	GGGGGCC	CTC ATC	CCTGCCA	TCTAC	ATGTT	r GGT	CTT	CCTC	
20	CTGGGCACCA	CGGGAAACGG	TCTGGTG	CTC TGG	ACCGTGT	TTCGG	AGCAG	CCC	GGA	GAAG	-186
	AGGCGCTCAG	CTGATATCTI	CATTGCT	AGC CTG	GCGGTGG	CTGAC	CTGAC	CTI	CGT	GGTG	240
	ACGCTGCCCC	TGTGGGCTAC	CTACACG	TAC CGG	GACTATG	ACTGG	СССТТ	TGG	GAC	CTTC	300
	TTCTGCAAGC	TCAGCAGCTA	CCTCATC	TTC GTC	AACATGT	ACGCC	AGCGI	CTI	CTG	CCTC	360
	ACCGGCCTCA	GCTTCGACCG	CTACCTG	GCC ATC	GTGAGGC	CAGTG	GCCAA	TGC	TCG	GCT G	420
25	AGGCTGCGGG	TCAGCGGGGC	CGTGGCC	ACG GCA	GTTCTTT	GGGTG	CTGGC	CGC	CCT	CCTG	480
	GCCATGCCTG	TCATGGTGTT	ACGCACC	ACC GGG	GACTTGG	AGAAC	ACCAC	TAA	GGT	GCAG	540
	TGCTACATGG	ACTACTCCAT	GGTGGCC	ACT GTG	AGCTCAG	AGTGG	GCCTG	GGA	GGT	GGC	600
	CTTGGGGTCT	CGTCCACCAC	CGTGGGC	TTT GTG	STGCCCT	TCACC	ATCAT	GCI	'GAC	CTGT	660
	TACTTCTTCA	TCGCCCAAAC	CATCGCT	GGC CAC	TTCCGCA	AGGAAG	CGCAT	CGA	.GGGC	CTG	720
30	CGGAAGCGGC	GCCGGCTTAA	GAGCATC.	ATC GTG	STGCTGG	TGGTG	ACCTT	TGC	CCTG	STGC	780
	TGGATGCCCT	ACCACCTGGT	GAAGACG	CTG TAC	ATGCTGG	GCAGC	CTGCT	GCA	.CTGG	SCCC	840
	TGTGACTTTG	ACCTCTTCCT	CATGAAC	ATC TTC	CCTACT	GCACCT	rgc ኔጥ	CVC	ርሞልና	יתייטי	900

	AACAGCTG	CC T	CAAC	CCT'	r - CC'	TCTA'	TGCC	TTT	rrcg:	ACC.	CCCG	CTTC	CG C	CAGG	CCTG	C 9	960
	ACCTCCAT	GC T	CTGC	rgtg	G CC	AGAG	CAGG	TGC	GCAG(GCA	CCTC	CCAC	AG C	AGCA	GTGG	3 10	020
	GAGAAGTC	AG C	CAGC'	FACT	TT(CGGG	GCAC	AGC	CAGG	GGC (CCGG	CCCC	AA C	ATCG	GCAA(3 1(080
	GGTGGAGA	AC A	GATG	CACG	A GA	AATC	CATC	ccc:	raca(GCC .	AGGA	GACC	ст т	GTGG:	rtga(2 13	L40
5	TAG															11	L43
	(199) IN	FORM	ATIO	N FOI	R SE	Q ID	NO:	198:					•				
10	(i,	(B) (C) (D)	LEI TYI STI	NGTH PE: 8 RANDI POLO	: 380 amino EDNE: EY: 1	0 am: o ac: SS: not:	ino a id relev	acid: vant				***	٠.				
	(x:	i) SI	EQUE	NCE I	DESCI	RIPT	ION:	SEQ	ID 1	NO:1	98:						
15	Met 1	Glu	Glu	Gly	Gly 5	Asp	Phe	Asp	Asn	Tyr 10	туг	Gly	Ala	Asp	Asn 15	Gln	
	Ser	Glu	Cys	Glu 20	Tyr	Thr	Asp	Trp	Lys 25	Ser	Ser	Gly	Ala	Leu 30	Ile	Pro	
	Ala	Ile	Tyr 35	Met	Leu	.Val	Phe	Leu 40	Leu	Gly	Thr	Thr	Gly 45	Asn	Gly	Leu	
20	Val	Leu 50	Trp	Thr	Val	Phe	Arg 55	Ser	Ser	Arg	Glu	Lys 60	Arg	Arg	Ser	Ala	
	Asp 65	Ile	Phe	Ile	Ala	Ser 70	Leu	Ala	Val	Ala	Asp 75	Leu	Thr	Phe	Val	Val 80	
25	Thr	Leu	Pro	Leu	Trp 85	Ala	Thr	Tyr	Thr	Tyr 90	Arg	Asp	Tyr	Asp	Trp 95	Pro	
	Phe	Gly		Phe 100	Phe	Cys	Lys	Leu	Ser 105	Ser	Tyr	Leu	Ile	Phe 110	Val	Asn	
	Met	Tyr	Ala 115	Ser	Val	Phe	Cys	Leu 120	Thr	Gly	Leu	Ser	Phe 125	Asp	Arg	Tyr	
30	Leu	Ala 130	Ile	Val	Arg	Pro	Val 135	Ala	Asn	Ala	Arg	Leu 140	Arg	Leu	Arg	Val	
	Ser 145	Gly	Ala	Val	Ala	Thr 150	Ala	Val	Leu	Trp	Val 155	Leu	Ala	Ala	Leu	Leu 160	
35	Ala	Met	Pro	Val	Met 165	Val	Leu	Arg	Thr	Thr 170	Gly	Asp	Leu	Glu	Asn 175	Thr	

-		Thr	. Lys	.Val	.Gln 180	Cys	.Tyr	Met	Asp	Tyr 185	Ser	Met	Val	Ala	Thr 190		Ser	
		Ser	Glu	Trp 195	Ala	Trp	Glu	Val	Gly 200	Leu	Gly	Val	Ser	Ser 205	Thr	Thr	Val	
5	, d	Gly	Phe 210	Val	Val	Pro	Phe	Thr 215	Ile	Met	Leu	Thr	Cys 220	Tyr	Phe	Phe	Ile	
		Ala 225		Thr	Ile	Ala	Gly 230	His	Phe	Arg	Lys	Glu 235	Arg	Ile	Glu	Gly	Leu 240	
10		Arg	Lys	Arg	Arg	Arg 245	Leu	Lys	Ser	Ile	Ile 250	Val	Val	Leu	Val	Val 255	Thr	
		Phe	Ala	Leu	Cys 260	Trp	Met	Pro	Tyr	His 265	Leu	Val	Lys	Thr	Leu 270	Tyr	Met	
		Leu	Gly	Ser 275	Leu	Leu	His	Trp	Pro 280	Cys	Asp	Phe		Leu 285	Phe	Leu	Met	
15		Asn	Ile 290	Phe	Pro	Tyr	Cys	Thr 295	Cys	Ile	Ser	Tyr	Val 300	Asn	Ser	Cys	Leu	
		Asn 305		Phe	Leu	Tyr	Ala 310	Phe	Phe	Asp	Pro	Arg 315	Phe	Arg	Gln	Ala	Cys 320	
20				Met		Cys 325	Cys	Gly	Gln	Ser	Arg 330	Cys	Ala	Gly	Thr	Ser 335	His	
		Ser	Ser	Ser	Gly 340	Glu	Lys	Ser	Ala	Ser 345	Tyr	Ser	Ser	Gly	His 350	Ser	Gln ;	
		Gly		Gly 355	Pro	Asn	Met		Lys 360	Gly	Gly	Glu		Met 365	His	Glu	Lys	
25		Ser	Ile 370	Pro	Tyr	Ser		Glu 375	Thr	Leu	Val	Val	Asp 380		·			
	(200)	INF	ORMA	TION	FOR	SEQ	ID	NO:1	99:									
30		(i)	(A) (B) (C)	UENC LEN TYP STR TOP	GTH: E: n: ANDE	111 ucle DNES	9 ba ic a S: s	se p cid ingl	S: airs					·	•			
		(ii) MO	LECU	LE T	YPE:	DNA	(ge	nomi	c)				-	-			•
		(xi) SE	QUEN	CE DI	ESCR:	IPTI(ON:	SEQ :	ID N	0:19:	9:						
35	ATGAA	CTAC	C CG	CTAA	CGCT	GGAZ	AATG	GAC (CTCG	AGAA	CC TO	GGAG(GACC	r GT	rctg	GGAA	60	כ
	CTGGA	CAGA	T TG	GACA	ACTA	TAAG	CGAC	ACC :	rccc:	rggt	GG A	TAAP	CATC	r cto	GCCC.	rgcc	120)

	ACAGAGGGTC CCCTCATGGC CTCCTTCAAG GCCGTGTTCG TGCCCGTGGC CTACAGCCTC	180
	ATCTTCCTCC TGGGCGTGAT CGGCAACGTC CTGGTGCTGG TGATCCTGGA GCGGCACCGG	240
	CAGACACGCA GTTCCACGGA GACCTTCCTG TTCCACCTGG CCGTGGCCGA CCTCCTGCTG	300
	GTCTTCATCT TGCCCTTTGC CGTGGCCGAG GGCTCTGTGG GCTGGGTCCT GGGGACCTTC	360
5	CTCTGCAAAA CTGTGATTGC CCTGCACAAA GTCAACTTCT ACTGCAGCAG CCTGCTCCTG	420
	GCCTGCATCG CCGTGGACCG CTACCTGGCC ATTGTCCACG CCGTCCATGC CTACCGCCAC	480
	CGCCGCCTCC TCTCCATCCA CATCACCTGT GGGACCATCT GGCTGGTGGG CTTCCTCCTT	540
	GCCTTGCCAG AGATTCTCTT CGCCAAAGTC AGCCAAGGCC ATCACAACAA CTCCCTGCCA	600
	CGTTGCACCT TCTCCCAAGA GAACCAAGCA GAAACGCATG CCTGGTTCAC CTCCCGATTC	660
10	CTCTACCATG TGGCGGGATT CCTGCTGCCC ATGCTGGTGA TGGGCTGGTG CTACGTGGGG	720
	GTAGTGCACA GGTTGCGCCA GGCCCAGCGG CGCCCTCAGC GGCAGAAGGC AAAAAGGGTG	780
	GCCATCCTGG TGACAAGCAT CTTCTTCCTC TGCTGGTCAC CCTACCACAT CGTCATCTTC	840
	CTGGACACCC TGGCGAGGCT GAAGGCCGTG GACAATACCT GCAAGCTGAA TGGCTCTCTC	900
	CCCGTGGCCA TCACCATGTG TGAGTTCCTG GGCCTGGCCC ACTGCTGCCT CAACCCCATG	960
15	CTCTACACTT TCGCCGGCGT GAAGTTCCGC AGTGACCTGT CGCGGCTCCT GACCAAGCTG	1020
	GGCTGTACCG GCCCTGCCTC CCTGTGCCAG CTCTTCCCTA GCTGGCGCAG GAGCAGTCTC	1080
	TCTGAGTCAG AGAATGCCAC CTCTCTCACC ACGTTCTAG	1119
	(201) INFORMATION FOR SEQ ID NO:200:	
20	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 372 amino acids	,
20	(B) TYPE: amino acid (C) STRANDEDNESS:	
	(D) TOPOLOGY: not relevant (ii) MOLECULE TYPE: protein	
25	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:200:	
	Met Asn Tyr Pro Leu Thr Leu Glu Met Asp Leu Glu Asn Leu Glu As 1 10 15	p
	Leu Phe Trp Glu Leu Asp Arg Leu Asp Asn Tyr Asn Asp Thr Ser Le	u
30	Val Glu Asn His Leu Cys Pro Ala Thr Glu Gly Pro Leu Met Ala Se 35 40 45	r

Phe Lys Ala Val Phe Val Pro Val Ala Tyr Ser Leu Ile Phe Leu Leu

	:	,	50 .		1,	• •		55	•		٠	, ,	60	٠			
٠		Gly 65	Val	Ile	Gly	Asn	Val 70	Leu	Val	Leu	Val	Ile 75	Leu	Glu	Arg	His	Arg 80
5		Gln	Thr	Arg	Ser	Ser 85	Thr	Glu	Thr	Phe	Leu 90	Phe	His	Leu	Ala	Val 95	Ala
		Asp	Leu	Leu	Leu 100	Val	Phe	Ile	Leu	Pro 105	Phe	Ala	Val	Ala	Glu 110	Gly	Ser
		Val	Gly	Trp 115	Val	Leu	Gly	Thr	Phe 120	Leu	Cys	Lys	Thr	Val 125	Ile	Ala	Let
10		His	Lys 130	Val	Asn	Phe	Tyr	Cys 135	Ser	Ser	Leu	Leu	Leu 140	Ala	Cys	Ile	Ala
	• .	Val 145	Asp	Arg	Tyr		Ala 150	Ile	Val	His	Ala	Val 155	His	Ala	Tyr	Arg	His 160
15		Arg	Arg	Leu	Leu	Ser 165	Ile	His	Ile	Thr	Cys 170	Gly	Thr	Ile	Trp	Leu 175	Val
		Gly	Phe	Leu	Leu 180		Leu	Pro	Glu	Ile 185	Leu	Phe	Ala	Lys	Val 190	Ser	Gln
-		Gly	His	His 195	Asn	Asn	Ser	Leu	Pro 200	Arg	Cys	Thr	Phe	Ser 205	Gln	Glu	Asn
20			Ala 210	Glu	Thr	His	Ala	Trp 215		Thr	Ser	Arg	Phe 220	Leu	Tyr	His	Val
		Ala 225	Gly	Phe	Leu	Leu	Pro 230	Met	Leu	Val	Met	Gly 235	Trp	Cys	Tyr	Val	Gly 240
25		Val	Val	His	Arg	Leu 245	Arg	Gln	Ala	Gln	Arg 250	Arg	Pro	Gln	Arg	Gln 255	Lys
		Ala	Lys	Arg	Val 260	Ala	Ile	Leu	Val	Thr 265	Ser	Ile	Phe	Phe	Leu 270	Cys	Trp
		Ser	Pro	Tyr 275	His	Ile	Val	Ile	Phe 280	Leu	Asp	Thr	Leu	Ala 285	Arg	Leu	Lys
30		Ala	Val 290	Asp	Asn	Thr	Cys	Lys 295	Leu	Asn	Gly	Ser	Leu 300	Pro	Val	Ala	Ile
	*	Thr 305	Met	Cys	Glu	Phe	Leu 310	Gly	Leu	Ala	His	Cys 315	Cys	Leu	Asn	Pro	Met 320
35		Leu	Tyr	Thr	Phe	Ala 325	Gly	Val	Lys	Phe	Arg 330	Ser	Asp	Leu	Ser	Arg 335	Leu
		Leu	Thr	Lys	Leu 340	Gly	Cys	Thr	Gly	Pro 345	Ala	Ser	Leu	Cys	Gln 350	Leu	Phe

Pro Ser Trp Arg Arg Ser Ser Leu Ser Glu Ser Glu Asn Ala Thr Ser 355 360 365

Leu Thr Thr Phe 370

- 5 (202) INFORMATION FOR SEQ ID NO:201:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1128 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
- 10 (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:201:

	ATGGATGTGA	CTTCCCAAGC	CCGGGGCGTG	GGCCTGGAGA	TGTACCCAGG	CACCGCGCAG	60
	CCTGCGGCCC	CCAACACCAC	CTCCCCGAG	CTCAACCTGT	CCCACCCGCT	CCTGGGCACC	120
15	GCCCTGGCCA	ATGGGACAGG	TGAGCTCTCG	GAGCACCAGC	AGTACGTGAT	CGGCCTGTTC	180
	CTCTCGTGCC	TCTACACCAT	CTTCCTCTTC	CCCATCGGCT	TTGTGGGCAA	CATCCTGATC	240
	CTGGTGGTGA	ACATCAGCTT	CCGCGAGAAG	ATGACCATCC	CCGACCTGTA	CTTCATCAAC	300
	CTGGCGGTGG	CGGACCTCAT	CCTGGTGGCC	GACTCCCTCA	TTGAGGTGTT	CAACCTGCAC	360
	GAGCGGTACT	ACGACATCGC	CGTCCTGTGC	ACCTTCATGT	CGCTCTTCCT	GCAGGTCAAC	420
20	ATGTACAGCA	GCGTCTTCTT	CCTCACCTGG	ATGAGCTTCG	ACCGCTACAT	CGCCCTGGCC	480
	AGGGCCATGC	GCTGCAGCCT	GTTCCGCACC	AAGCACCACG	CCCGGCTGAG	CTGTGGCCTC	540
•	ATCTGGATGG	CATCCGTGTC	AGCCACGCTG	GTGCCCTTCA	CCGCCGTGCA	CCTGCAGCAC	600
	ACCGACGAGG	CCTGCTTCTG	TTTCGCGGAT	GTCCGGGAGG	TGCAGTGGCT	CGAGGTCACG	660
	CTGGGCTTCA	TCGTGCCCTT	CGCCATCATC	GGCCTGTGCT	ACTCCCTCAT	TGTCCGGGTG	720
25	CTGGTCAGGG	CGCACCGGCA	CCGTGGGCTG	CGGCCCCGGC	GGCAGAAGGC	GAAGCGCATG	780
	ATCCTCGCGG	TGGTGCTGGT	CTTCTTCGTC	TGCTGGCTGC	CGGAGAACGT	CTTCATCAGC	840
	GTGCACCTCC	TGCAGCGGAC	GCAGCCTGGG	GCCGCTCCCT	GCAAGCAGTC	TTTCCGCCAT	900
	GCCCACCCC	TCACGGGCCA	CATTGTCAAC	CTCACCGCCT	TCTCCAACAG	CTGCCTAAAC	960
	CCCCTCATCT	ACAGCTTTCT	CGGGGAGACC	TTCAGGGACA	AGCTGAGGCT	GTACATTGAG	1020
30	CAGAAAACAA	ATTTGCCGGC	CCTGAACCGC	TTCTGTCACG	CTGCCCTGAA	GGCCGTCATT	1080
	CCAGACAGCA	CCGAGCAGTC	GGATGTGAGG	TTCAGCAGTG	CCGTGTAG		1128

(203) INFORMATION FOR SEQ ID NO:202: ...

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 375 amino acids
 - (B) TYPE: amino acid
- (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:202: Met Asp Val Thr Ser Gln Ala Arg Gly Val Gly Leu Glu Met Tyr Pro 10 Gly Thr Ala Gln Pro Ala Ala Pro Asn Thr Thr Ser Pro Glu Leu Asn 25 Leu Ser His Pro Leu Leu Gly Thr Ala Leu Ala Asn Gly Thr Gly Glu 35 35 40 15 Leu Ser Glu His Gln Gln Tyr Val Ile Gly Leu Phe Leu Ser Cys Leu 50 55 Tyr Thr Ile Phe Leu Phe Pro Ile Gly Phe Val Gly Asn Ile Leu Ile 65 70 75 Leu Val Val Asn Ile Ser Phe Arg Glu Lys Met Thr Ile Pro Asp Leu 20 9.0 Tyr Phe Ile Asn Leu Ala Val Ala Asp Leu Ile Leu Val Ala Asp Ser 100 105 Leu Ile Glu Val Phe Asn Leu His Glu Arg Tyr Tyr Asp Ile Ala Val 120 25
- Leu Cys Thr Phe Met Ser Leu Phe Leu Gln Val Asn Met Tyr Ser Ser 130 135 140
 - Val Phe Phe Leu Thr Trp Met Ser Phe Asp Arg Tyr Ile Ala Leu Ala 145 150 155 160
- Arg Ala Met Arg Cys Ser Leu Phe Arg Thr Lys His His Ala Arg Leu 30 165 170 175
 - Ser Cys Gly Leu Ile Trp Met Ala Ser Val Ser Ala Thr Leu Val Pro 180 185 190
 - Phe Thr Ala Val His Leu Gln His Thr Asp Glu Ala Cys Phe Cys Phe 195 200 205
- Ala Asp Val Arg Glu Val Gln Trp Leu Glu Val Thr Leu Gly Phe Ile
 210 215 220
 - Val Pro Phe Ala Ile Ile Gly Leu Cys Tyr Ser Leu Ile Val Arg Val

	225		. •	230					235			• •	•	240	
	Leu Va	al Arg		is Arg 15	His	Arg	Gly	Leu 250	Arg	Pro	Arg	Arg	Gln 255	Lys	
5	Ala Ly	ys Arg	Met Il 260	le Leu	Ala	Val	Val 265	Leu	Val	Phe	Phe	Val 270	Cys	Trp	
	Leu Pr	ro Glu 275	Asn Va	al Phe	Ile	Ser 280	Val	His	Leu	Leu	Gln 285	Arg	Thr	Gln	
		ly Ala 90	Ala P	ro Cys	Lys 295	Gln	Ser	Phe	Arg		Ala	His	Pro	Leu	
10	Thr G:	ly His	Ile Va	al Asn 310	Leu	Thr	Ala	Phe	Ser 315	Asn	Ser	Cys	Leu	Asn 320	
	Pro L	eu Ile	Tyr S				Glu					Lys		Arg	
15	Leu.T	yr Ile	Glu G	ln Lys	Thr	Asn	Leu 345	Pro	Ala	Leu	Asn	Arg 350	Phe	Cys	
	His A	la Ala 355	Leu L	ys Ala	Val	Ile 360		Asp	Ser	Thr	Glu 365		Ser	Asp	
		rg Phe 70	Ser S	er Ala	Val 375		,				. •	*,		-	
20	(204) INFO	RMATIO	N FOR	SEQ ID	NO:	203:									
	(i)	SEQUENCE (A) LE	NGTH: PE: nu	1137 b cleic	ase : acid	pair	s ·								
25		(C) ST				le			• •		•	*.			
٠	(ii)	MOLEC	ULE TY	PE: DN	A (g	enom	ic)				٠	• .			
	(xi)	SEQUE	NCE DE	SCRIPT	'ION:	SEQ	ID	NO:2	03:						
	ATGGACCTGG	GGAAA	CCAAT	GAAAAG	CGTG	CTG	GTGG	TGG	CTCT	CCTT	GT C	ATTT	TCCA	G	60
	GTATGCCTGT	GTCAA	GATGA	GGTCAC	:GGAC	GAT	TACA	TCG	GAGA	CAAC	AC C	ACAG	TGGA	C	120
30	TACACTTTGT	TCGAG	TCTTT	GTGCTC	CAAG	AAG	GACG	TGC	GGAA	CTTT	AA A	GCCI	GGTT	c	180
	CTCCCTATCA	TGTAC	TCCAT	CATTTO	TTTC	GTG	GGCC	TAC	TGGG	CAAT	GG G	CTGG	TCGT	G	240
	TTGACCTATA	A TCTAT	TTCAA	GAGGCT	CAAG	ACC	ATGA	CCG	ATAC	CTAC	CT G	CTCA	ACCI	'G	300
	GCGGTGGCAG	ACATO	CTCTT	CCTCCI	GACC	CTI	ccci	TCT	GGGC	CTAC	AG C	CGCGG	CCAA	.G	360
	TCCTGGGTCT	r TCGGI	GTCCA	CTTTTC	CAAG	CTC	ATCI	TTG	CCAT	CTAC	AA G	ATG	GCTI	c	420

	TTCAGTGGCA	TGCTCCTACT	TCTTTGCATC	AGCATTGACC	GCTACGTGGC	CÁTCGTCCAG	480
	GCTGTCTCAG	CTCACCGCCA	CCGTGCCCGC	GTCCTTCTCA	TCAGCAAGCT	GTCCTGTGTG	540
	GGCATCTGGA	TACTAGCCAC	AGTGCTCTCC	ATCCCAGAGC	TCCTGTACAG	TGACCTCCAG	600
	AGGAGCAGCA	GTGAGCAAGC	GATGCGATGC	TCTCTCATCA	CAGAGCATGT	GGAGGCCTTT	660
5	ATCACCATCC	AGGTGGCCCA	GATGGTGATC	GGCTTTCTGG	TCCCCCTGCT	GGCCATGAGC	720
	TTCTGTTACC	TTGTCATCAT	CCGCACCCTG	CTCCAGGCAC	GCAACTTTGA	GCGCAACAAG	780
	GCCAAAAAGG	TGATCATCGC	TGTGGTCGTG	GTCTTCATAG	TCTTCCAGCT	GCCCTACAAT	840
	GGGGTGGTCC	TGGCCCAGAC	GGTGGCCAAC	TTCAACATCA	CCAGTAGCAC	CTGTGAGCTC	900
	AGTAAGCAAC	TCAACATCGC	CTACGACGTC	ACCTACAGCC	TGGCCTGCGT	CCGCTGCTGC	960
10	GTCAACCCTT	TCTTGTACGC	CTŢCATCGGC	GTCAAGTTCC	GCAACGATCT	CTTCAAGCTC	1020
	TTCAAGGACC	TGGGCTGCCT	CAGCCAGGAG	CAGCTCCGGC	AGTGGTCTTC	CTGTCGGCAC	1080
	ATCCGGCGCT	CCTCCATGAG	TGTGGAGGCC	GAGACCACCA	CCACCTTCTC	CCCATAG	1137
	(205) INFO	RMATION FOR	SEQ ID NO:2	204:			

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 378 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein
- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:204:

Met Asp Leu Gly Lys Pro Met Lys Ser Val Leu Val Val Ala Leu Leu 1 5 10 15

Val Ile Phe Gln Val Cys Leu Cys Gln Asp Glu Val Thr Asp Asp Tyr
20 25 30

25 Ile Gly Asp Asn Thr Thr Val Asp Tyr Thr Leu Phe Glu Ser Leu Cys
35 40 45

Ser Lys Lys Asp Val Arg Asn Phe Lys Ala Trp Phe Leu Pro Ile Met 50 55 60

Tyr Ser Ile Ile Cys Phe Val Gly Leu Leu Gly Asn Gly Leu Val Val 30 65 70 75 80

Leu Thr Tyr Ile Tyr Phe Lys Arg Leu Lys Thr Met Thr Asp Thr Tyr 85 . 90 95

Leu Leu Asn Leu Ala Val Ala Asp Ile Leu Phe Leu Leu Thr Leu Pro

		•		100				•	105				٠.٠	110		
	Phe	Trp	Ala 115	Tyr	Ser	Ala	Ala	Lys 120	Ser	Trp	Val	Phe	Gly 125	Val	His	Phe
5	Cys	Lys 130	Leu	Ile	Phe	Ala	Ile 135	Tyr	Lys	Met	Ser	Phe 140	Phe	Ser	Gly	Met
	Leu 145	Leu	Leu	Leu	Cys	Ile 150	Ser	Ile	Asp	Arg	Tyr 155	Val	.Ala	Ile	Val	Gln 160
	Ala	Val	Ser	Ala	His 165	Arg	His	Arg	Ala	Arg 170	Val	Leu	Leu	Ile	Ser 175	Lys
10	Leu	Ser	Cys	Val 180	Gly	Ile	Trp	Ile	Leu 185	Ala	Thr	Val	Leu	Ser 190	Ile	Pro
	Glu	Leu	Leu 195	Tyr	Ser	Asp		Gln 200	Arg	Ser	Ser	Ser	Glu 205		Ala	Met
15	Arg	Cys 210	Ser	Leu	Ile	Thr	Glu 215	His	Val	Glu	Ala	Phe 220	Ile	Thr	Ile	Gln
	Val 225	Ala	Gln	Met	Val	Ile 230	Gly	Phe	Leu	Val	Pro 235	Leu	Leu	Ala	Met	Ser 240
	Phe	Cys	Tyr	Leu	Val 245	Ile	Ile	Arg	Thr	Leu 250	Leu	Gln	Ala	Arg	Asn 255	Phe
20	Glu	Arg	Asn	Lys 260	Ala	Lys	Lys	Val	Ile 265	Ile	Ala	Val	Val	Val 270	Val	Phe
	Ile	Val	Phe 275	Gln	Leu	Pro	Tyr	Asn 280		Val	Val	Leu	Ala 285	Gln	Thr	Val
25	Ala	Asn 290		Asn	Ile	Thr	Ser 295	Ser	Thr	Cys	Glu	Leu 300		Lys	Gln	Leu
	Asn 305		Ala	Tyr	Asp	Val 310		Tyr	Ser	Leu	Ala 315		Val	Arg	Cys	Cys 320
	Val	Asn	Pro	Phe	Leu 325		Ala	Phe	Ile	Gly 330		Lys	Phe	Arg	Asn 335	Asp
30	Leu	Phe	Lys	Leu 340		Lys	Asp	Leu	Gly 345		Leu	Ser	Gln	Glu 350		Leu
	Arg	Gln	355	Ser	Ser	Cys	Arg	His 360		Arg	Arg	Ser	Ser 365		Ser	Val
35	Glu	Ala 370		Thr	Thr	Thr	Thr 375		Ser	Pro	•					

(206) INFORMATION FOR SEQ ID NO:205:

PRICESSID JAIO MOSSORS IA

WO 00/22129 PCT/US99/23938

161

- (A) LENGTH: 1086 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5

30

- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:205:

ATGGATATAC AAATGGCAAA CAATTTTACT CCGCCCTCTG CAACTCCTCA GGGAAATGAC 60 TGTGACCTCT ATGCACATCA CAGCACGGCC AGGATAGTAA TGCCTCTGCA TTACAGCCTC 120 10 GTCTTCATCA TTGGGCTCGT GGGAAACTTA CTAGCCTTGG TCGTCATTGT TCAAAACAGG 180 AAAAAAATCA ACTCTACCAC CCTCTATTCA ACAAATTTGG TGATTTCTGA TATACTTTTT 240 ACCACGGCTT TGCCTACACG AATAGCCTAC TATGCAATGG GCTTTGACTG GAGAATCGGA 300 GATGCCTTGT GTAGGATAAC TGCGCTAGTG TTTTACATCA ACACATATGC AGGTGTGAAC . 360 TTTATGACCT GCCTGAGTAT TGACCGCTTC ATTGCTGTGG TGCACCCTCT ACGCTACAAC 420 AAGATAAAAA GGATTGAACA TGCAAAAGGC GTGTGCATAT TTGTCTGGAT TCTAGTATTT 480 GCTCAGACAC TCCCACTCCT CATCAACCCT ATGTCAAAGC AGGAGGCTGA AAGGATTACA 540 TGCATGGAGT ATCCAAACTT TGAAGAAACT AAATCTCTTC CCTGGATTCT GCTTGGGGCA 600 TGTTTCATAG GATATGTACT TCCACTTATA ATCATTCTCA TCTGCTATTC TCAGATCTGC 660 TGCAAACTCT TCAGAACTGC CAAACAAAAC CCACTCACTG AGAAATCTGG TGTAAACAAA 720 20. AAGGCTAAAA ACACAATTAT TCTTATTATT GTTGTGTTTG TTCTCTGTTT CACACCTTAC 780 CATGTTGCAA TTATTCAACA TATGATTAAG AAGCTTCGTT TCTCTAATTT CCTGGAATGT 840 AGCCAAAGAC ATTCGTTCCA GATTTCTCTG CACTTTACAG TATGCCTGAT GAACTTCAAT 900 TGCTGCATGG ACCCTTTTAT CTACTTCTTT GCATGTAAAG GGTATAAGAG AAAGGTTATG 960 AGGATGCTGA AACGGCAAGT CAGTGTATCG ATTTCTAGTG CTGTGAAGTC AGCCCCTGAA 1020 25 GAAAATTCAC GTGAAATGAC AGAAACGCAG ATGATGATAC ATTCCAAGTC TTCAAATGGA 1080 AAGTGA 1086

(207) INFORMATION FOR SEQ ID NO:206:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 361 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant

	(ii	L) MC	LEC	JLE 7	TYPE:	pro	oteir	ı	-		·· .	•				
	(x)	i) SE	EQUE	NCE I	DESCI	RIPTI	ON:	SEQ	ID 1	10:20	6:					
	Met 1	Asp	Ile	Gln	Met 5	Ala	Asn	Asn	Pho	Thr 10	Pro	Pro	Ser	Ala	Thr 15	Pro
5	Gln	Gly	Asn	Asp 20	Cys	Asp	Leu	Tyr	Ala 25	His	His	Ser	Thr	Ala 30	Arg	Ile
	Val	Met	Pro 35	Leu	His	Tyr	Ser	Leu 40	Val	Phe	Ile	Ile	Gly 45	Leu	Val	Gly
10	Asn	Leu 50	Leu	Ala	Leu	Val	Val 55	Ile	Val	Gln	Asn	Arg 60	Lys	Lys	Ile	Asn
	Ser 65	Thr	Thr	Leu	Tyr	Ser 70	Thr	Asn	Leu		75		Asp			80
	Thr	Thr	Ala	Leu	Pro 85	Thr	Arg	Ile	Ala							
15	Trp	Arg	Ile	Gly 100	Asp	Ala	Leu	Cys	Arg 105	Ile	Thr	Ala	Leu	Val 110	Phe	Tyr
	Ile	Asn	Thr 115	Tyr	Ala	Gly	Val		Phe			Cys	Leu 125		Ile	Asp
20	Arg	Phe 130	Ile	Ala	Val	Val	His 135	Pro	Leu	Arg	Tyr	Asn 140	Lys	Ile	Lys	Arg
	Ile 145	Glu	His	Ala	Lys	Gly 150	-		Ile		Val 155	Trp	Ile		Val	Phe 160
	Ala	Gln	Thr	Leu	Pro 165	Leu	Leu	Ile	Asn	Pro 170	Met	Ser	Lys	Gln	Glu 175	Ala
25	Glu	Arg	Ile	Thr 180	Cys	Met	Glu	Tyr	Pro 185	Asn	Phe	Glu		Thr 190		Ser
	Leu	Pro	195	Ile	Leu	Leu	Gly	Ala 200	Cys	Phe	Ile	Gly	205		Leu	
30	Leu	Ile 210			Leu	Ile	Cys 215		Ser	Gln	Ile	Cys 220	Cys	Lys	Leu	Phe
•	Arg 225	Thr	Ala	Lys	Gln	Asn 230	Pro	Leu	Thr	Glu	Lys 235	Ser	Gly	Val	Asn	Lys 240
	Lys	Ala	Lys	Asn	Thr 245	Ile	Ile	Leu	Ile	Ile 250	Val	Val	Phe	Val	Leu 255	Cys
35	Phe	Thr	Pro	Tyr 260	His	Val	Ala	Ile	Ile 265	Gln	His	Met	Ile	Lys 270	Lys	Leu

	Arg	Phe	Ser 275	Asn	Phe	Leu	Glu	-Cys 280	Ser	Gln	Arg	His	Ser 285	Phe	Gln	Ile	
	Ser	Leu 290		Phe	Thr	Val	Cys 295	Leu	Met	Asn	Phe	Asn 300	Cys	Cys	Met	Asp	
5	Pro 305		Ile	Tyr	Phe	Phe 310	Ala	Cys	Lys	Gly	Tyr 315	Lys	Arg	Lys	Val	Met 320	
	Arg	Met	Leu	Lys	Arg 325	Gln	Val	Ser	Val	Ser 330	Ile	Ser	Ser	Ala	Val 335	Lys	
10	Ser	Ala	Pro	Glu 340	Glu	Asn	Ser	Arg	Glu 345	Met	Thr	Glu	Thr	Gln 350	Met	Met	
	Ile	His	Ser 355	Lys	Ser	Ser		Gly 360	Lys								•
	(208) IN	FORM	OITA	1 FOR	SEÇ) ID	NO:2	07:									
15	(i)	(A) (B) (C)	LEN TYI STR	CE CH IGTH: PE: n RANDE	144 ucle DNES	6 ba ic a SS: s	se p cid singl	airs								· .	
	(ii	L) MC	LECU	LE I	YPE:	DNA	ı (ge	nomi	.c)								
20	. (xi	L) SE	EQUEN	ICE _, E	ESCR	IPTI	ON:	SEQ	ID N	10:20	7:					A.	
	ATGCGGTGG	C TO	TGGC	CCCT	. GGC	TGTC	TCT	CTTG	CTGT	GA I	TTTG	GCTG	T GG	GGCT	'AAGC	!	60
	AGGGTCTCT	rg ge	GGTG	CCCC	CCI	GCAC	CTG	GGCA	.GGCA	CA G	AGCC	GAGA	.c cc	AGGA	.GCAG	: : 1	20
	CAGAGCCGA	AT CC	AAGA	\GGGG	CAC	CGAG	GAT	GAGG	AGGC	CA A	GGGC	GTGC	A GC	'AGTA	TGTG	1	80
	CCTGAGGAG	ST GG	GCGG	SAGTA	ccc	ccgg	ccc	ATTC	ACCC	TG C	TGGC	CTGC	A GC	CAAC	CAAG	2	40
25	CCCTTGGTG	G CC	ACCA	'GCCC	TAA	.cccc	GAC	AAGG	ATGG	IGG G	CACC	CCAG	A CA	GTGG	GCAG	3	00
	GAACTGAGG	G GC	'AATC	TGAC	AGG	GGCA	CCA	GGGC	AGAG	GC I	ACAG	ATCC	A GA	ACCC	CCTG	3	60
	TATCCGGTG	A CC	GAGA	GCTC	CTA	CAGT	GCC	TATG	CCAT	CA I	GCTT	CTGG	C GC	TGGT	GGTG	4:	20
	TTTGCGGTG	G GC	ATTG	TGGG	CAA	CCTG	TCG	GTCA	TGTG	CA T	CGTG	TGGC	A CA	GCTA	CTAC	4	80
	CTGAAGAGC	G CC	TGGA	ACTC	CAT	CCTT	GCC	AGCC	TGGC	CC I	CTGG	GATT	T TC	TGGT	CCTC	54	40
30	TTTTTCTGC	C TC	CCTA	TTGT	CAT	CTTC	AAC	GAGA	TCAC	CA A	GCAG	AGGC'	T AC	TGGG	TGAC	6	00
	GTTTCTTGT	'C GT	GCCG	TGCC	CTT	CATG	GAG	GTCT	CCTC	TC T	GGGA	GTCA	C GA	CTTT	CAGC	60	60
	CTCTGTGCC	C TG	GGCA	TTGA	CCG	CTTC	CAC	GTGG	CCAC	CA G	CACC	CTGC	C CA	AGGT	GAGG	7:	20
	CCCATCGAG	C GG	TGCC	א אדר	СРД	ССТС	GCC	<u>አ</u> ልሮጥ	ייננר	ጥር ጥ	ראיזרי	тссс	T GG	CCTC	ሮልሞሮ	7.	20

	ACGCTGGCTG	TGCCTGAGCT	CCTGCTGTGG	CAGCTGGCAC	AGGAGCCTGC	CCCCACCATG	840
	GGCACCCTGG	ACTCATGCAT	CATGAAACCC	TCAGCCAGCC	TGCCCGAGTC	CCTGTATTCA	900
	CTGGTGATGA	CCTACCAGAA	CGCCCGCATG	TGGTGGTACT	TTGGCTGCTA	CTTCTGCCTG	960
	CCCATCCTCT	TCACAGTCAC	CTGCCAGCTG	GTGACATGGC	GGGTGCGAGG	CCCTCCAGGG	1020
5	AGGAAGTCAG	AGTGCAGGGC	CAGCAAGCAC	GAGCAGTGTG	AGAGCCAGCT	CAAGAGCACC	1080
	GTGGTGGCC	TGACCGTGGT	CTACGCCTTC	TGCACCCTCC	CAGAGAACGT	CTGCAACATC	1140
	GTGGTGGCCT	ACCTCTCCAC	CGAGCTGACC	CGCCAGACCC	TGGACCTCCT	GGGCCTCATC	1200
	AACCAGTTCT	CCACCTTCTT	CAAGGGCGCC	ATCACCCCAG	TGCTGCTCCT	TTGCATCTGC	1260
	AGGCCGCTGG	GCCAGGCCTT	CCTGGACTGC	TGCTGCTGCT	GCTGCTGTGA	GGAGTGCGGC	1320
10	GGGGCTTCGG	AGGCCTCTGC	TGCCAATGGG	TCGGACAACA	AGCTCAAGAC	CGAGGTGTCC	1380
	TCTTCCATCT	ACTTCCACAA	GCCCAGGGAG	TCACCCCCAC	TCCTGCCCCT	GGGCACACCT	1440
	TGCTGA						1446

(209) INFORMATION FOR SEQ ID NO:208:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 481 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein
- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:208:

Met Arg Trp Leu Trp Pro Leu Ala Val Ser Leu Ala Val Ile Leu Ala 1 5 10 15

Val Gly Leu Ser Arg Val Ser Gly Gly Ala Pro Leu His Leu Gly Arg 20 25 30

25 His Arg Ala Glu Thr Gln Glu Gln Gln Ser Arg Ser Lys Arg Gly Thr 35 40 45

Glu Asp Glu Glu Ala Lys Gly Val Gln Gln Tyr Val Pro Glu Glu Trp
50 55 60

Ala Glu Tyr Pro Arg Pro Ile His Pro Ala Gly Leu Gln Pro Thr Lys 65 70 75 80

Pro Leu Val Ala Thr Ser Pro Asn Pro Asp Lys Asp Gly Gly Thr Pro 85 90 95

Asp Ser Gly Gln Glu Leu Arg Gly Asn Leu Thr Gly Ala Pro Gly Gln

		,			٠.,	100	. :-				105		•	;		110	•	
			Arg	Leu	Gln 115	Ile	Gln	Asn	Pro	Leu 120	Tyr	Pro	Val	Thr	Glu 125	Ser	Ser	Tyr
5			Ser	Ala 130	Tyr	Ala	Ile	Met	Leu 135	Leu	Ala	Leu	Val	Val 140	Phe	Ala	Val	Gly
		٠	Ile 145	Val	Gly	Asn	Leu	Ser 150	Val	Met	Cys	Ile	Val 155	Trp	His	Ser	Tyr	Tyr 160
	•		Leu	Lys	Ser	Ala	Trp 165	Asn	Ser	Ile	Leu	Ala 170	Ser	Leu	Ala	Leu	Trp 175	Asp
10	ē		Phe	Leu	Val	Leu 180	Phe	Phe	Cys	Leu	Pro 185	Ile	Val	Ile	Phe	Asn 190	Glu	Ile
			Thr	Lys	Gln 195	Arg	Leu	Leu	Gly	Asp 200	Val	Ser	Cys	Arg	Ala 205	Val	Pro	Phe
15			Met	Glu 210	Val	Ser	Ser	Leu	Gly 215	Val	Thr	Thr	Phe	Ser 220	Leu	`Cys	Ala	Leu
			Gly 225	Ile	Asp	Arg	Phe	His 230	Val	Ala	Thr	Ser	Thr 235	Leu	Pro	Lys		Arg 240
			Pro	Ile	Glu	Arg	Cys 245	Gln	Ser	Ile	Leu	Ala 250	Lys	Leu	Ala	Val	Ile. 255	Trp
20			Val	Gly	Ser	Met 260	Thr	Leu	Ala		Pro 265	Glu	Leu	Leu	Leu	Trp 270	Gln	Leu
			Ala	Gln	Glu 275	Pro	Ala	Pro	Thr	Met 280	Gly	Thr	Leu	Asp	Ser 285	Cys	Ile	Met
25		٠,	Lys	Pro 290	Ser	Ala	Ser	Leu	Pro 295	Glu	Ser	Leu	Tyr		Leu		Met	Thr
			Tyr 305	Gln	Asn	Ala	Arg	Met 310	Trp	Trp	Tyr	Phe	Gly 315	Cys	Tyr	Phe	Cys	Leu 320
		٠		Ile		Phe	Thr 325	Val	Thr	Cys	Gln	Leu 330	Val	Thr	Trp	Arg	Val 335	Arg
30			Gly	Pro	Pro	Gly 340	Arg	Lys	Ser	Glu	Cys 345	Arg	Ala	Ser	Lys	His 350	Glu	Gln
			Cys	Glu	Ser 355		Leu	Lys	Ser	Thr 360	Val	Val	Gly	Leu	Thr 365	Val	Val	Tyr
35		-	Ala	Phe 370	Cys	Thr	Leu	Pro	Glu 375	Asn	Val	Cys	Asn	Ile 380	Val	Val	Ala	Tyr
			Leu 385	Ser	Thr	Glu	Leu	Thr 390		Gln	Thr	Leu ·	Asp 395	Leu	Leu	Gly	Leu	Ile 400

	Ash Gin Phe Ser Thr Phe Phe Lys Gly Ala He Thr Pro Val Leu Leu 405 410 415	u
٠	Leu Cys Ile Cys Arg Pro Leu Gly Gln Ala Phe Leu Asp Cys Cys Cys 420 425 430	S
5	Cys Cys Cys Cys Glu Glu Cys Gly Gly Ala Ser Glu Ala Ser Ala Ala 435 440 445	a
	Asn Gly Ser Asp Asn Lys Leu Lys Thr Glu Val Ser Ser Ser Ile Typ 450 455 460	r
10	Phe His Lys Pro Arg Glu Ser Pro Pro Leu Leu Pro Leu Gly Thr Pro	
	Cys	
	(210) INFORMATION FOR SEQ ID NO:209:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1101 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:209:	
	ATGTGGAACG CGACGCCCAG CGAAGAGCCG GGGTTCAACC TCACACTGGC CGACCTGGAC	60
	TGGGATGCTT CCCCCGGCAA CGACTCGCTG GGCGACGAGC TGCTGCAGCT CTTCCCCGCG	120
	CCGCTGCTGG CGGGCGTCAC AGCCACCTGC GTGGCACTCT TCGTGGTGGG TATCGCTGGC	180
	AACCTGCTCA CCATGCTGGT GGTGTCGCGC TTCCGCGAGC TGCGCACCAC CACCAACCTC	240
25	TACCTGTCCA GCATGGCCTT CTCCGATCTG CTCATCTTCC TCTGCATGCC CCTGGACCTC	300
	GTTCGCCTCT GGCAGTACCG GCCCTGGAAC TTCGGCGACC TCCTCTGCAA ACTCTTCCAA	360
	TTCGTCAGTG AGAGCTGCAC CTACGCCACG GTGCTCACCA TCACAGCGCT GAGCGTCGAG	420
	CGCTACTTCG CCATCTGCTT CCCACTCCGG GCCAAGGTGG TGGTCACCAA GGGGCGGGTG	480
	AAGCTGGTCA TCTTCGTCAT CTGGGCCGTG GCCTTCTGCA GCGCCGGGCC CATCTTCGTG	540
30	CTAGTCGGGG TGGAGCACGA GAACGGCACC GACCCTTGGG ACACCAACGA GTGCCGCCCC	600
	ACCGAGTTTG CGGTGCGCTC TGGACTGCTC ACGGTCATGG TGTGGGTGTC CAGCATCTTC	660
	TTCTTCCTTC CTGTCTTCTG TCTCACGGTC CTCTACAGTC TCATCGGCAG GAAGCTGTGG	720

CGGAGGAGGC GCGGCGATGC TGTCGTGGGT GCCTCGCTCA GGGACCAGAA CCACAAGCAA 780

שובטרוני אוט מוסטומנו ווא

	ACCAAGA	L AAA	GCTG	GCTG	T AG	TGGT	GTTI	GCC	TTCA	TCC	TCTG	CTGG	CT C	CCCI	TCCA	'C	840
	GTAGGGC	GAT A	ATTT.	TTTT	C CA	AATC	CTTT	' GAG	CCTG	GCT	CCTT	'GGAG	TA	GCTC	AGAT	C	900
	AGCCAGT	ACT G	CAAC	CTCG	T GT	CCTT	TGTC	CTC	TTCT	ACC	TCAG	TGCI	GC C	'ATCA	ACCC	C	960
	ATTCTGT	ACA A	CATC	ATGT	C CA	AGAA	GTAC	CGG	GTGG	CAG	TĠTT	CAGA	CT T	'CTGG	GATT	C 1	020
5	GAACCCT	rcr c	CCAG	AGAA	A GC	TCTC	CACT	CTG	AAAG	ATG	AAAG	TTCT	cg g	GCCT	GGAC	A 1	080
	GAATCTAG	TA T	TAAT	ACAT	G A											1	101
	(211) In	IFORM	ATIO	n Fo	R SE	Q ID	NO:	210:									
10	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 366 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: (D) TOPOLOGY: not relevant (ii) MOLECULE TYPE: protein																
						-	•				. •			•			
1.5		i) S															
15	Met 1	Trp	Asn	Ala	Thr 5	Pro	Ser	Glu	Glu	Pro 10	Gly	Phe	Asn	Leu	Thr 15	Leu	,
	Ala	Asp	Leu	Asp 20	Trp	Asp	Ala	Ser	Pro 25	Gly	Asn	Asp	Ser	Leu 30	Gly	Asp	
20	Glu	Leu	Leu 35	Gln	Leu	Phe	Pro	Ala 40	Pro	Leu	Leu	Ala	Gly 45	Val	Thr	Ala	}e4
	Thr	Cys 50	Val	Ala	Leu		Val 55	Val			Ala	_	Asn	Leu	Leu	Thr	
	Met 65	Leu	Val	Val	Ser	Arg 70	Phe	Arg	Glu	Leu	Arg 75	Thr	Thr	Thr	Asn [.]	Leu 80	
25	Tyr	Leu	Ser	Ser	Met 85	Ala	Phe	Ser	Asp	Leu 90	Leu	Ile	Phe		Cys 95	Met	
	Pro	Leu	Asp	Leu 100	Val	Arg	Leu	Trp	Gln 105	Tyr	Arg	Pro	Trp	Asn 110	Phe	Gly	
30	Asp	Leu	Leu 115	Cys	Lys	Leu	Phe	Gln 120	Phe	Val	Ser	Glu	Ser 125	Суз	Thr	Tyr	
	Ala	Thr 130	Val	Leu	Thr	Ile	Thr 135	Ala	Leu	Ser	Val	Glu 140	Arg	Tyr	Phe	Ala	
	Ile 145	Cys	Phe	Pro	Leu	Arg 150	Ala	Lys	Val	Val	Val 155	Thr	Lys	Gly	Arg	Val 160	
35 .	Lys	Leu	Val	Ile	Phe	Val	Ile	Trp	Ala	Val	Ala	Phe	Cys	Ser	Ala	Gly	

					165	:	:		2	1.70					175		
	Pro	Ile	Phe	Val 180		Val	Gly	Val	Glu 185		Glu	Asn	Gly	Thr 190	Asp	Pro	
5	Trp	Asp	Thr 195		Glu	Cys	Arg	Pro 200	Thr	Glu	Phe	Ala	Val 205	Arg	Ser	Gly	
	Leu	Leu 210	Thr	Val	Met	Val	Trp 215	Val	Ser	Ser	Ile	Phe 220	Phe	Phe	Leu	Pro	
	Val 225	Phe	Cys	Leu	Thr	Val 230	Leu	Tyr	Ser	Leu	Ile 235	Gly	Arg	Lys	Leu	Trp 240	
10	Arg	Arg	Arg	Arg	Gly 245	Asp	Ala	Val	Val	Gly 250	Ala	Ser	Leu	Arg	Asp 255	Gln	
	Asn	His	Lys	Gln 260		Lys	Lys	Met	Leu 265	Ala	Val	Val		Phe 270	Ala	Phe	
15	Ile	Leu	Cys 275	Trp	Leu	Pro	Phe	His 280	Val	Gly	Arg	Tyr	Leu 285	Phe	Ser	Lys	
-	Ser	Phe 290	Glu	Pro	Gly	Ser	Leu 295	Glu	Ile	Ala	Gln	Ile 300	Ser	Gln	Tyr	Cys	
	Asn 305	Leu	Val	Ser	Phe	Val 310	Leu	Phe	Tyr	Leu	Ser 315	Ala	Ala	Ile	Asn	Pro 320	
20	Ile	Leu	Tyr	Asn	Ile 325	Met	Ser	Lys	Lys	Tyr 330	Arg	Val	Ala	Val	Phe 335	-	
:	Leu	Leu	Gly	Phe 340	Glu	Pro	Phe	Ser	Gln 345	Arg	Lys	Leu	Ser	Thr 350	Leu	Lys	
25	Asp		Ser 355	Ser	Arg	Ala	Trp	Thr 360	Glu	Ser	Ser	Ile	Asn 365		· · · ·		
	(212) IN	FORMA	TION	FOR	SEÇ) ID	NO:2	11:									
	· (i)		LEN	IGTH :	184	2 ba	se p		i								
30		(C)	STR	PE: n CANDE POLOG	DNES	S: s	ingl	e									
	(ii	.) MC	LECU	ĻE, T	YPE:	DNA	(ge	nomi	c)			٠			·		
	(xi	.) SE	QUEŅ	CE,D	ESCR	IPTI	ON:	SEQ	ID N	0:21	1.:		. ,				
	ATGCGAGCC	C CG	GGCG	CGCT	TCT	CGCC	CGC	ATGT	CGCG	GC T	ACTG	CTTC	T GC	TACT	GCTC	•	50
35	AAGGTGTCT	G CC	TCTT	CTGC	CCT	CGGG	GTC	GCCC	CTGC	GT C	CAGA	AACG.	A AA	CTTG	TCTG	12	20
	GGGGAGAGC	T GT	GCAC	CTAC	AGŢ	GATC	CAG	CGCC	GCGG	CA G	GGAC	GCCT	G GG	GACC	GGGA	18	30

	AATTCTGCAA	GAGACGITCI	GCGAGCCCGA	GCACCCAGGG	AGGAGCAGGG	GGCAGCGIII	240
	CTTGCGGGAC	CCTCCTGGGA	CCTGCCGGCG	GCCCCGGGCC	GTGACCCGGC	TGCAGGCAGA	300
	GGGGCGGAGG	CGTCGGCAGC	CGGACCCCCG	GGACCTCCAA	CCAGGCCACC	TGGCCCCTGG	360
	AGGTGGAAAG	GTGCTCGGGG	TCAGGAGCCT	TCTGAAACTT	TGGGGAGAGG	GAACCCCACG	420
5	GCCCTCCAGC	TCTTCCTTCA	GATCTCAGAG	GAGGAAGAGA	AGGGTCCCAG	AGGCGCTGGC	480
	ATTTCCGGGC	GTAGCCAGGA	GCAGAGTGTG	AAGACAGTCC	CCGGAGCCAG	CGATCTTTTT	540
	TACTGGCCAA	GGAGAGCCGG	GAAACTCCAG	GGTTCCCACC	ACAAGCCCCT	GTCCAAGACG	600
	GCCAATGGAC	TGGCGGGGCA	CGAAGGGTGG	ACAATTGCAC	TCCCGGGCCG	GGCGCTGGCC	660
	CAGAATGGAT	CCTTGGGTGA	AGGAATCCAT	GAGCCTGGGG	GTCCCCGCCG	GGGAAACAGC	720
0	ACGAACCGGC	GTGTGAGACT	GAAGAACCCC	TTCTACCCGC	TGACCCAGGA	GTCCTATGGA	780
	GCCTACGCGG	TCATGTGTCT	GTCCGTGGTG	ATCTTCGGGA	CCGGCATCAT	TGGCAACCTG	· 840
	GCGGTGATGT	GCATCGTGTG	CCACAACTAC	TACATGCGGA	GCATCTCCAA	CTCCCTCTTG	900
٠.	GCCAACCTGG	CCTTCTGGGA	CTTTCTCATC	ATCTTCTTCT	GCCTTCCGCT	GGTCATCTTC	960
	CACGAGCTGA	CCAAGAAGTG	GCTGCTGGAG	GACTTCTCCT	GCAAGATCGT	GCCCTATATA	1020
5	GAGGTCGCCT	CTCTGGGAGT	CACCACTTTC	ACCTTATGTG	CTCTGTGCAT	AGACCGCTTC	1080
	CGTGCTGCCA	CCAACGTACA	GATGTACTAC	GAAATGATCG	AAAATTGTTC	CTCAACAACT	1140
	GCCAAACTTG	CTGTTATATG	GGTGGGAGCT	CTATTGTTAG	CACTTCCAGA	AGTTGTTCTC	1200
	CGCCAGCTGA	GCAAGGAGGA	TTTGGGGTTT	AGTGGCCGAG	CTCCGGCAGA	AAGGTGCATT	1260
	ATTAAGATCT	CTCCTGATTT	ACCAGACACC	ATCTATGTTC	TAGCCCTCAC	CTACGACAGT	1320
20	GCGAGACTGT	GGTGGTATTT	TGGCTGTTAC	TTTTGTTTGC	CCACGCTTTT	CACCATCACC	1380
	TGCTCTCTAG	TGACTGCGAG	GAAAATCCGC	AAAGCAGAGA	AAGCCTGTAC	CCGAGGGAAT	1440
	AAACGGCAGA	TTCAACTAGA	GAGTCAGATG	AAGTGTACAG	TAGTGGCACT	GACCATTTTA	1500
	TATGGATTTT	GCATTATTCC	TGAAAATATC	TGCAACATTG	TTACTGCCTA	CATGGCTACA	1560
	GGGGTTTCAC	AGCAGACAAT	GGACCTCCTT	AATATCATCA	GCCAGTTCCT	TTTGTTCTTT	1620
25	AAGTCCTGTG	TCACCCCAGT	CCTCCTTTTC	TGTCTCTGCA	AACCCTTCAG	TCGGGCCTTC	1680
	ATGGAGTGCT	GCTGCTGTTG	CTGTGAGGAA	TGCATTCAGA	AGTCTTCAAC	GGTGACCAGT	1740
	GATGACAATG	ACAACGAGTA	CACCACGGAA	CTCGAACTCT	CGCCTTTCAG	TACCATACGC	1800
	CGTGAAATGT	CCACTTTTGC	TTCTGTCGGA	ACTCATTGCT	GA		184

	(213) IN	FORM	OITA	1 FOI	R SEQ	Q ID	NO:2	212:	5.							
5	(i)	(B)	QUENC LEI TYI STI	NGTH: PE: & RANDI	: 613 amino EDNES	3 ami o aci SS:	ino a id	acids	3				•			
	(1:	i) M(OLEC	JLE T	TYPE:	: pro	oteir	ı								
	(x:	i) SI	EQUE	VCE I	DESCI	RIPTI	ON:	SEQ	ID 1	NO:21	12:					
10	Met 1	Arg	Ala	Pro	Gly 5	Ala	Leu	Leu	Ala	Arg 10	Met	Ser	Arg	Leu	Leu 15	Leu
	Leu	Leu	Leu	Leu 20	Lys	Val	Ser	Ala	Ser 25	Ser	Ala	Leu	Gly	Val 30	Ala	Pro
	- ··Ala	Ser	Arg 35	Asn	Glu	Thr	Cys	Leu 40	Gly	Glu	Ser	Cys	Ala 45	Pro	Thr	Val
15	Ile	Gln 50	Arg	Arg	Gly	Arg	Asp 55	Ala	Trp	Gly	Pro	Gly 60	Asn	Ser	Ala	Arg
	Asp 65	Val	Leu	Arg	Ala	Arg 70	Ala	Pro	Arg	Glu	Glu 75	Gln	Gly	Ala	Ala	Phe 80
20	Leu	Ala	Gly	Pro	Ser 85	Trp	Asp	Leu	Pro	Ala 90	Ala	Pro	Gly	Arg	Asp 95	Pro
	Ala	Ala	Gly	Arg 100	Gly	Ala	Glu	Ala	Ser 105	Ala	Ala	Gly	Pro	Pro	Gly	Pro
	Pro	Thr	Arg 115	Pro	Pro	Gly	Pro	Trp 120	Arg	Trp	Lys	Gly	Ala 125	Arg	Gly	Gln
25	Glu	Pro 130	Ser	Glu	Thr	Leu	Gly 135	Arg	Gly	Asn	Pro	Thr 140	Ala	Leu	Gln	Leu
	Phe 145	Leu	Gln	Ile	Ser	Glu 150	Glu	Glu	Glu	Lys	Gly 155	Pro	Arg	Gly	Ala	Gly 160
30	Ile	Ser	Gly	Arg	Ser 165	Gln	Glu	Gln	Ser	Val 170	Lys	Thr	Val	Pro	Gly 175	Ala
	Ser	Asp	Leu	Phe 180	Tyr	Trp	Pro	Arg	Arg 185	Ala	Gly	Lys	Leu	Gln 190	Gly	Ser
	His	His	Lys 195	Pro	Leu	Ser	Lys	Thr 200	Ala	Asn	Gly	Leu	Ala 205	Gly	His	Glu
35	Gly	Trp 210	Thr	Ile	Ala	Leu	Pro 215	Gly	Arg	Ala	Leu	Ala 220	Gln	Asn	Gly	Ser
																•

Leu Gly Glu Gly Ile His Glu Pro Gly Gly Pro Arg Arg Gly Asn Ser

שוכטטטום אווט מואזי מוסטיוטניאין ואי

Thr Asn Arg Arg Val Arg Leu Lys Asn Pro Phe Tyr Pro Let 245 Glu Ser Tyr Gly Ala Tyr Ala Val Met Cys Leu Ser Val Va 265 Gly Thr Gly Ile Ile Gly Asn Leu Ala Val Met Cys Ile Va 285 Asn Tyr Tyr Met Arg Ser Ile Ser Asn Ser Leu Leu Ala As 300 Phe Trp Asp Phe Leu Ile Ile Phe Phe Cys Leu Pro Leu Va 305 His Glu Leu Thr Lys Lys Trp Leu Leu Glu Asp Phe Ser Cy 325 Val Pro Tyr Ile Glu Val Ala Ser Leu Gly Val Thr Thr Ph 340 Cys Ala Leu Cys Ile Asp Arg Phe Arg Ala Ala Thr Asn Va 355 Tyr Tyr Glu Met Ile Glu Asn Cys Ser Ser Thr Thr Ala Ly 370 Val Ile Trp Val Gly Ala Leu Leu Leu Ala Leu Pro Glu Va 385 Arg Gln Leu Ser Lys Glu Asp Leu Gly Phe Ser Gly Arg Ala Ala Glu Arg Ala Ala Cys Ala Leu Ser Lys Glu Asp Leu Gly Phe Ser Gly Arg Ala Ala Glu Val Ala Ser Leu Gly Phe Ser Gly Arg Ala Ala Glu Val Ala Ser Leu Leu Ala Leu Pro Glu Va 385 Arg Gln Leu Ser Lys Glu Asp Leu Gly Phe Ser Gly Arg Ala Ala Glu Arg Ala	255 l Ile Phe 0 l Cys His
Gly Thr Gly Ile Ile Gly Asn Leu Ala Val Met Cys Ile Va 275 Asn Tyr Tyr Met Arg Ser Ile Ser Asn Ser Leu Leu Ala As 300 Phe Trp Asp Phe Leu Ile Ile Phe Phe Cys Leu Pro Leu Va 305 His Glu Leu Thr Lys Lys Trp Leu Leu Glu Asp Phe Ser Cy 325 Val Pro Tyr Ile Glu Val Ala Ser Leu Gly Val Thr Thr Ph 340 Cys Ala Leu Cys Ile Asp Arg Phe Arg Ala Ala Thr Asn Va 355 Tyr Tyr Glu Met Ile Glu Asn Cys Ser Ser Thr Thr Ala Ly 370 Val Ile Trp Val Gly Ala Leu Leu Leu Ala Leu Pro Glu Va 385 Arg Gln Leu Ser Lys Glu Asp Leu Gly Phe Ser Gly Arg Ala 405	0 l Cys His n Leu Ala l Ile Phe
Asn Tyr Tyr Met Arg Ser Ile Ser Asn Ser Leu Leu Ala As 300 Phe Trp Asp Phe Leu Ile Ile Phe Phe Cys Leu Pro Leu Va 305 His Glu Leu Thr Lys Lys Trp Leu Leu Glu Asp Phe Ser Cy 325 Val Pro Tyr Ile Glu Val Ala Ser Leu Gly Val Thr Thr Ph 340 Cys Ala Leu Cys Ile Asp Arg Phe Arg Ala Ala Thr Asn Va 355 Tyr Tyr Glu Met Ile Glu Asn Cys Ser Ser Thr Thr Ala Ly 370 Val Ile Trp Val Gly Ala Leu Leu Leu Ala Leu Pro Glu Va 385 Arg Gln Leu Ser Lys Glu Asp Leu Gly Phe Ser Gly Arg Ala Arg	n Leu Ala
290 295 300 Phe Trp Asp Phe Leu Ile Ile Phe Phe Cys Leu Pro Leu Va 315 His Glu Leu Thr Lys Lys Trp Leu Leu Glu Asp Phe Ser Cy 325 Val Pro Tyr Ile Glu Val Ala Ser Leu Gly Val Thr Thr Ph 345 Cys Ala Leu Cys Ile Asp Arg Phe Arg Ala Ala Thr Asn Va 355 Tyr Tyr Glu Met Ile Glu Asn Cys Ser Ser Thr Thr Ala Ly 370 Val Ile Trp Val Gly Ala Leu Leu Leu Ala Leu Pro Glu Va 385 Arg Gln Leu Ser Lys Glu Asp Leu Gly Phe Ser Gly Arg Al 405	l Ile Phe
305 His Glu Leu Thr Lys Lys Trp Leu Leu Glu Asp Phe Ser Cy 325 Val Pro Tyr Ile Glu Val Ala Ser Leu Gly Val Thr Thr Ph 340 Cys Ala Leu Cys Ile Asp Arg Phe Arg Ala Ala Thr Asn Va 355 Tyr Tyr Glu Met Ile Glu Asn Cys Ser Ser Thr Thr Ala Ly 370 Val Ile Trp Val Gly Ala Leu Leu Leu Ala Leu Pro Glu Va 385 Arg Gln Leu Ser Lys Glu Asp Leu Gly Phe Ser Gly Arg Al 405	
Val Pro Tyr Ile Glu Val Ala Ser Leu Gly Val Thr Thr Ph 340 Cys Ala Leu Cys Ile Asp Arg Phe Arg Ala Ala Thr Asn Va 355 Tyr Tyr Glu Met Ile Glu Asn Cys Ser Ser Thr Thr Ala Ly 370 Val Ile Trp Val Gly Ala Leu Leu Leu Ala Leu Pro Glu Va 385 Arg Gln Leu Ser Lys Glu Asp Leu Gly Phe Ser Gly Arg Al 405	
Cys Ala Leu Cys Ile Asp Arg Phe Arg Ala Ala Thr Asn Va 355 Tyr Tyr Glu Met Ile Glu Asn Cys Ser Ser Thr Thr Ala Ly 370 Val Ile Trp Val Gly Ala Leu Leu Leu Ala Leu Pro Glu Va 385 Arg Gln Leu Ser Lys Glu Asp Leu Gly Phe Ser Gly Arg Al 405	s Lys Ile 335
Tyr Tyr Glu Met Ile Glu Asn Cys Ser Ser Thr Thr Ala Ly 370 375 380 Val Ile Trp Val Gly Ala Leu Leu Leu Ala Leu Pro Glu Va 385 390 395 Arg Gln Leu Ser Lys Glu Asp Leu Gly Phe Ser Gly Arg Al 405 410	
20 Val Ile Trp Val Gly Ala Leu Leu Leu Ala Leu Pro Glu Va 385 390 395 Arg Gln Leu Ser Lys Glu Asp Leu Gly Phe Ser Gly Arg Al 405 410	l Gln Met
385 390 395 Arg Gln Leu Ser Lys Glu Asp Leu Gly Phe Ser Gly Arg Al 405 410	s Leu Ala
405 410	l Val Leu 400
	a Pro Ala 415
Glu Arg Cys Ile Ile Lys Ile Ser Pro Asp Leu Pro Asp Th 420 425 43	_
Val Leu Ala Leu Thr Tyr Asp Ser Ala Arg Leu Trp Trp Ty 435 440 445	: Phe Gly
Cys Tyr Phe Cys Leu Pro Thr Leu Phe Thr Ile Thr Cys Se 450 455 460	Leu Val
Thr Ala Arg Lys Ile Arg Lys Ala Glu Lys Ala Cys Thr Arg 465 470 475	g Gly Asn 480
Lys Arg Gln Ile Gln Leu Glu Ser Gln Met Lys Cys Thr Va 485 490	Val Ala 495
Leu Thr Ile Leu Tyr Gly Phe Cys Ile Ile Pro Glu Asn Ile 500 505 516	
Ile Val Thr Ala Tyr Met Ala Thr Gly Val Ser Gln Gln Thr	Met Asp

		Leu Asn 530	Ile.	Ile	Ser	Gln 535	Phe	Leu	Leu	Phe	Phe 540	Lys	Ser	Cys	Val	
٠.	Thr 1 545	Pro Val	Leu		Phe 550	Cys	Leu	Cys	Lys	Pro 555	Phe	Ser	Arg	Ala	Phe 560	
5	Met (Glu Cys		Cys 565	Cys	Cys	Cys	Glu	Glu 570	Cys	Ile	Gln	Lys	Ser 575	Ser	
٠.	Thr \	Val Thr	Ser . 580	Asp	Asp	Asn	Asp	Asn 585	Glu	Tyr	Thr	Thr	Glu 590	Leu	Glu	
10	Leu S	Ser Pro	Phe :	Ser	Thr	Ile	Arg 600	Arg	Glu	Met	Ser	Thr 605	Phe	Ala	Ser	
• •		Sly Thr	His	Cys												
	(214) INFO	RMATION	FOR	SEQ	ID:	NO:2	13:-			٠.						
15	(i)	SEQUENC	GTH:	124	8 ba	se p		٠.		٠.		٠.				
		(B) TYP (C) STR (D) TOP	ANDEI	DNES	S: s	ingl	.e				:					
	(ii)	MOLECU	LE T	YPE:	DNA	· (ge	nomi	c) -				<i>3</i> ~				
20	(xi)	SEQUEN	CE DI	ESCR	IPTI	ON:	SEQ	ID N	0:21	3:						
	ATGGTTTTTG	CTCACA	GAAT	GGA	TAAC	AGC	AAGC	CACA	TT T	GATT	ATTC	C TA	CACT	TCTG		5
	GTGCCCCTCC	AAAACC	GCAG	CTG	CACT	GAA	ACAG	CCAC	AC C	TCTG	CCAA	G CC	AATA	.CCTG	12	2 (
	ATGGAATTAA	GTGAGG.	AGCA	CAG'	TTGG	ATG	AGCA	ACCA	AA C	AGAC	CTTC	A CT	ATGT	GCTG	18	3 (
	AAACCCGGGG	AAGTGG	CCAC	AGC	CAGC	ATC	TTCT	TTGG	GA T	TCTG	TGGT	T GT	TTTC	TATC	24	1 (
25	TTCGGCAATT	CCCTGG	TTTG	TTT	GGTC	ATC	CATA	GGAG'	TA G	GAGG	ACTC.	A GT	CTAC	CACC	3 ()(
	AACTACTTTG	TGGTCT	CCAT	GGC	ATGT	GCT	GACC	TTCT	CA T	CAGC	GTTG	C CA	GCAC	GCCT	36	5 (
	TTCGTCCTGC	TCCAGT	TCAC	CAC'	rggaj	AGG	TGGA	CGCT	GG G	TAGT	GCAA	C GT	GCAA	GGTT	42	? (
	GTGCGATATT	TTCAAT	ATCT	CAC	rccad	GGT (GTCC.	AGAT	CT A	CGTT	CTCC	т ст	CCAT	CTGC	48	3 (
	ATAGACCGGT	TCTACA	CCAT	CGT	CTATO	CCT	CTGA	GCTT	CA A	GGTG	TCCA	G AG	AAAA	AGCC	54	1
30	AAGAAAATGA	TTGCGG	CATC	GTG	GATC	TTT (GATG(CAGG	CT T	IGTG.	ACCC	C TG	TGCT	CTTT	60) (
	TTCTATGGCT	CCAACT	GGGA	CAG	CATT	rgt i	AACT	ATTT	CC T	cccc'	rccr(C TT	GGGA	AGGC	66	; (
	ACTGCCTACA	CTGTCA:	rcca.	CTT	CTTGC	etg (GGCT'	TTGT	GA T	rcca'	rctg:	r cc	rcat:	AATT	72	: C
	TTATTTTACC	AAAAGG'	CAT	AAAZ	TAT	י ידיר	TGGA	ידעענ	AG G	מממר	SATC	2 CC	מאבי	ZCTC	70	

	AGGAGGACAA TGAACATTGT CCCTCGGACA AAAGTGAAAA CTAAAAAGAT GTTCCTCATT 840
	TTAAATCTGT TGTTTTTGCT CTCCTGGCTG CCTTTTCATG TAGCTCAGCT ATGGCACCCC 900
	CATGAACAAG ACTATAAGAA AAGTTCCCTT GTTTTCACAG CTATCACATG GATATCCTTT 960
	AGTTCTTCAG CCTCTAAACC TACTCTGTAT TCAATTTATA ATGCCAATTT TCGGAGAGGG 1020
5	ATGAAAGAGA CTTTTTGCAT GTCCTCTATG AAATGTTACC GAAGCAATGC CTATACTATC 1080
	ACAACAAGTT CAAGGATGGC CAAAAAAAAC TACGTTGGCA TTTCAGAAAT CCCTTCCATG 1140
	GCCAAAACTA TTACCAAAGA CTCGATCTAT GACTCATTTG ACAGAGAAGC CAAGGAAAAA 1200
	AAGCTTGCTT GGCCCATTAA CTCAAATCCA CCAAATACTT TTGTCTAA 1248
	(215) INFORMATION FOR SEQ ID NO:214:
10	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 415 amino acids
	(B) TYPE: amino acid (C) STRANDEDNESS:
	(D) TOPOLOGY: not relevant
15	(ii) MOLECULE TYPE: protein
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:214:
	Met Val Phe Ala His Arg Met Asp Asn Ser Lys Pro His Leu Ile Ile: 1 5 10 15
20	Pro Thr Leu Leu Val Pro Leu Gln Asn Arg Ser Cys Thr Glu Thr Ala 20 25 30
	Thr Pro Leu Pro Ser Gln Tyr Leu Met Glu Leu Ser Glu Glu His Ser 35 40 45
	Trp Met Ser Asn Gln Thr Asp Leu His Tyr Val Leu Lys Pro Gly Glu 50 55 60
25	Val Ala Thr Ala Ser Ile Phe Phe Gly Ile Leu Trp Leu Phe Ser Ile 65 70 75 80
	Phe Gly Asn Ser Leu Val Cys Leu Val Ile His Arg Ser Arg Arg Thr 85 90 95
30	Gln Ser Thr Thr Asn Tyr Phe Val Val Ser Met Ala Cys Ala Asp Leu 100 105 110
	Leu Ile Ser Val Ala Ser Thr Pro Phe Val Leu Leu Gln Phe Thr Thr 115 120 125
	Gly Arg Trp Thr Leu Gly Ser Ala Thr Cys Lys Val Val Arg Tyr Phe 130 135 140

	Gln 145	Tyr	Leu	Thr	Pro	Gly 150	Val	Gln	Ile	Tyr	Val 155	Leu	Leu	Ser	Ile	Суя 160
	Ile	Asp	Arg	Phe	Tyr 165	Thr	Ile	Val	Tyr	Pro 170	Leu	Ser	Phe	Lys	Val 175	Ser
5	Arg	Glu	Lys	Ala 180	Lys	Lys	Met	Ile	Ala 185	Ala	Ser	Trp	Ile	Phe 190	Asp	Ala
	Gly	Phe	Val 195	Thr	Pro	Val	Leu	Phe 200	Phe	Tyr	Gly	Ser	Asn 205	Trp	Asp	Ser
10	His	Cys 210	Asn	Tyr	Phe	Leu	Pro 215	Ser	Ser	Trp	Glu	Gly 220	Thr	Ala	Tyr	Thr
	Val 225	Ile	His	Phe	Leu	Val 230	Gly	Phe	Val	Ile	Pro 235	Ser	Val	Leu	Ile	Ile 240
	Leu	Phe	Tyr	Gln	Lys 245	Val	Ile	Lys	Tyr	Ile 250	_	Arg	Ile	Gly	Thr 255	Asp
15	Gly	Arg	Thr	Val 260	Arg	Arg	Thr	Met	Asn 265	Ile	Val	Pro	Arg	Thr 270	Lys	Val
	Lys	Thr	Lys 275	Lys	Met	Phe	Leu	Ile 280	Leu	Asn	Leu	Leu	Phe 285	Leu	Leu	Ser
20	Trp	Leu 290	Pro	Phe	His	Val	Ala 295		Leu	Trp	His	Pro 300	His	Glu	Gln	Asp
	Туг 305	Lys	Lys	Ser	Ser	Leu 310	Val	Phe	Thr	Ala	Ile 315	Thr	Trp	Ile	Ser	Phe 320
	Ser 	Ser	Ser	Ala	Ser 325	Lys	Pro	Thr	Leu	Tyr 330	Ser	Ile	Tyr		Ala 335	Asn
25	Phe	Arg	Arg	Gly 340	Met	Lys	Glu	Thr	Phe 345	Cys	Met	Ser	Ser	Met 350	Lys	Cys
	Tyr	Arg	Ser 355	Asn	Ala	Tyr	Thr	Ile 360	Thr	Thr	Ser	Ser	Arg 365	Met	Ala	Lys
30	Lys	Asn 370	Tyr	Val	Gly	Ile	Ser 375	Glu	Ile	Pro	Ser	Met 380	Ala	Lys	Thr	Ile
	Thr 385		Asp								Arg 395		Ala	Lys		Lys 400
	Lys	Leu	Ala	Trp	Pro 405	Ile	Asn	Ser	Asn	Pro 410	Pro	Asn	Thr	Phe	Val 415	
35	(216) IN	FORM	ATION	1 FOF	SEC] ID	NO:2	215:								

(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1842 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- 5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:215:

·	ATGGGGCCCA	CCCTAGCGGT	TCCCACCCCC	TATGGCTGTA	TTGGCTGTA	GCTACCCCAG	60
	CCAGAATACC	CACCGGCTCT	AATCATCTTT	ATGTTCTGCG	CGATGGTTAT	CACCATCGTT	120
	GTAGACCTAA	TCGGCAACTC	CATGGTCATT	TTGGCTGTGA	CGAAGAACAA	GAAGCTCCGG	180
:	AATTCTGGCA	ACATCTTCGT	GGTCAGTCTC	TCTGTGGCCG	ATATGCTGGT	GGCCATCTAC	240
10	CCATACCCTT	TGATGCTGCA	TGCCATGTCC	ATTGGGGGCT	GGGATCTGAG	CCAGTTACAG	300
.4	TGCCAGATGG	TCGGGTTCAT	CACAGGGCTG	AGTGTGGTCG	GCTCCATCTT	CAACATCGTG	360
	GCAATCGCTA	TCAACCGTTA	CTGCTACATC	TGCCACAGCC	TCCAGTACGA	ACGGATCTTC	420
v	AGTGTGCGCA	ATACCTGCAT	CTACCTGGTC	ATCACCTGGA	TCATGACCGT	CCTGGCTGTC	480
r'o	CTGCCCAACA	TGTACATTGG	CACCATCGAG	TACGATCCTC	GCACCTACAC	CTGCATCTTC	≟ 540
15	AACTATCTGA	ACAACCCTGT	CTTCACTGTT	ACCATCGTCT	GCATCCACTT	CGTCCTCCCT	600
	CTCCTCATCG	TGGGTTTCTG	CTACGTGAGG	ATCTGGACCA	AAGTGCTGGC	GGCCCGTGAC	660
	CCTGCAGGGC	AGAATCCTGA	CAACCAACTT	GCTGAGGTTC	GCAATAAACT	AACCATGTTT	- 72 0
	GTGATCTTCC	TCCTCTTTGC	AGTGTGCTGG	TGCCCTATCA	ACGTGCTCAC	TGTCTTGGTG	780
	GCTGTCAGTC	CGAAGGAGAT	GGCAGGCAAG	ATCCCCAACT	GGCTTTATCT	TGCAGCCTAC	840
20	TTCATAGCCT	ACTTCAACAG	CTGCCTCAAC	GCTGTGATCT	ACGGGCTCCT	CAATGAGAAT	900
	TTCCGAAGAG	AATACTGGAC	CATCTTCCAT	GCTATGCGGC	ACCCTATCAT	ATTCTTCTCT	960
	GGCCTCATCA	GTGATATTCG	TGAGATGCAG	GAGGCCCGTA	CCCTGGCCCG	CGCCCGTGCC	1020
	CATGCTCGCG	ACCAAGCTCG	TGAACAAGAC	CGTGCCCATG	CCTGTCCTGC	TGTGGAGGAA	1080
	ACCCCGATGA	ATGTCCGGAA	TGTTCCATTA	CCTGGTGATG	CTGCAGCTGG	CCACCCGAC	1140
25	CGTGCCTCTG	GCCACCCTAA	GCCCCATTCC	AGATCCTCCT	CTGCCTATCG	CAAATCTGCC	1200
	TCTACCCACC	ACAAGTCTGT	CTTTAGCCAC	TCCAAGGCTG	CCTCTGGTCA	CCTCAAGCCT	1260
	GTCTCTGGCC	ACTCCAAGCC	TGCCTCTGGT	CACCCCAAGT	CTGCCACTGT	CTACCCTAAG	1320
	CCTGCCTCTG	TCCATTTCAA	GGCTGACTCT	GTCCATTTCA	AGGGTGACTC	TGTCCATTTC	1380
	AAGCCTGACT	CTGTTCATTT	CAAGCCTGCT	TCCAGCAACC	CCAAGCCCAT	CACTGGCCAC	1440

	CATG	rctc:	rg c	rggc	AGCC	A CT	CCAA	GTCT	GCC'	TTCA	ATG	CTGC	CACC.	AG C	CACC	CTAA	A 1	500
	CCCA.	CAAC	GC C	AGCT	ACCA	G CC	ATGC'	TGAG	CCC	ACCA	CTG (CTGA	CTAT	cc c	AAGC	CTGC	C 1	560
	ACTA	CCAG	CC A	CCCT	AAGC	C CG	CTGC'	TGCT	GAC	AACC	CTG :	AGCT	CTCT	GC C	TCCC	ATTG	C 1	€20
	CCCG	AGATO	בכ, כי	rgccz	ATTG	c cci	ACCC'	TGTG	TCT	GACG	ACA (GTGA	CCTC	CC TO	GAGT	CGGC	С 1	680
5	TCTA	3CCC1	rg co	CGCT	GGC	CA	CCAA	GCCT	GCT	GCCA(GCC 2	AGCT	ggag'	rc To	GACA	CCAT	C 1	740
	GCTG	ACCTI	rc c	rgac(CCTA	TG:	TAGT	CACT	ACC	AGTA	CCA I	ATGA'	TTAC	CA TO	GATG'	rcgr(G 1	800
	GTTG	rtga:	rg T	rgaa(GATG	A TC	CTGA'	TGAA	ATG	GCTG:	rgt (GA					1	842
	(217)	INI	FORM	ATIOI	N FOI	SE(Q ID	NO:2	216:									
10	. • • • • •	(i)	(A) (B) (C) (D)	LEI TYI STI	NGTH PE: 8 RANDI POLO	: 61: amino EDNE: EY: 1	3 am: o ac: SS: not:	ino a id relev	acid: vant				·	"				
16.							•	oteir				.						
15		(x:	·	-					~									
			Gly	Pro	Thr	Leu 5	Ala	Val	Pro	Thr	Pro 10	Tyr	Gly	Cys	Ile	Gly 15	Cys	
i		Lys	Leu	Pro	Gln 20	Pro	Glu	Tyr	Pro		Ala		Ile	Ile	Phe	Met	Phe	
20		Cys	Ala	Met 35	Val	Ile	Thr	Ile	Val 40	Val	Asp	Leu	Ile	Gly 45	Asn	Ser	Met	
	٠.	Val			Ala		Thr	Lys 55		Lys		Leu		Asn	Ser	Gly	Asn	
25		Ile 65	Phe	Val	Val	Ser	Leu 70	Ser	Val	Ala	Asp	Met 75	Leu	Val	Ala	Ile	Tyr 80	
		Pro	Tyr	Pro	Leu	Met 85	Leu	His	Ala	Met	Ser 90	Ile	Gly	Gly	Trp	Asp 95	Leu	
	1 - 41	Ser	Gln	Leu	Gln 100	Cys	Gln	Met	Val	Gly 105	Phe	Ile	Thr	Gly	Leu 110	Ser	Val	
30		Val	Gly	Ser 115		Phe	Asn	Ile		Ala			Ile		Arg	Tyr	Cys	
		Tyr	Ile 130	Cys	His	Ser	Leu	Gln 135	Tyr	Glu	Arg	Ile	Phe 140	Ser	Val	Arg	Asn	
35:-		Thr 145	Cys	Ile	Tyr	Leu	Val 150	Ile	Thr	Trp	Ile	Met 155		Val	Leu	Ala	Val 160	

DESCRIPTION OF THE PROPERTY OF

:	• .	Leu	Pro	Asn	Met	Tyr 165	Ile	Gly	Thr	Ile	Glu 170	туг	Asp	Pro	Arg	Thr 175	Tyr
		Thr	Cys	Ile	Phe 180	Asn	Tyr	Leu	Asn	Asn 185	Pro	Val	Phe	Thr	Val 190	Thr	Ile
5		Val	Cys	Ile 195	His	Phe	Val	Leu	Pro 200	Leu	Leu	Ile	Val	Gly 205	Phe	Cys	Tyr
		Val	Arg 210	Ile	Trp	Thr	Lys	Val 215	Leu	Ala	Ala	Arg	Asp 220	Pro	Ala	Gly	Gln
10		Asn 225	Pro	Asp	Asn	Gln	Leu 230	Ala	Glu	Val	Arg	Asn 235	Lys	Leu	Thr	Met	Phe 240
		Val	Ile	Phe	Leu	Leu 245	Phe	Ala	Val	Cys	Trp 250	Cys	Pro	Ile	Asn	Val 255	Leu
				Leu	260					265					270		
15			_	Leu 275					280					285	• ,		4.
	* •		290	,	**			295	٠.				300				Glu
20		305	_	Thr			310				•	315					320
	•	-		Ile		325					330					335	
				Arg	340					345				·	350		
25				Cys 355					360					365			
			370					375					380				Gly
30	. •	385		Lys			390					395			,		400
				His		405					410		-			415	
				Lys	420					425					430		
35		-		Ala 435					440					445			
		Asp	Ser	Val	His	Phe	Lys	Gly	Asp	Ser	Val	His	Phe	Lys	Pro	Asp	Ser

			450					455					460					
		Val 465	His	Phe	Lys	Pro	Ala 470	Ser	Ser	Asn	Pro	Lys 475	Pro	Ile	Thr	Gly	His 480	
5		His	Val	Ser	Ala	Gly 485	Ser	His	Ser	Lys	Ser 490	Ala	Phe	Asn	Ala	Ala 495	Thr	
		Ser	His	Pro	Lys 500	Pro	Ile	Lys	Pro	Ala 505	Thr	Ser	His	Ala	Glu 510	Pro	Thr	
		Thr	Ala	Asp 515	Tyr	Pro	Lys	Pro	Ala 520	Thr	Thr	Ser	His	Pro 525	Lys	Pro	Ala	
10		Ala	Ala 530	Asp	Asn	Pro	Glu	Leu 535	Ser	Ala	Ser	His	Cys 540	Pro	Glu	Ile	Pro	
		Ala 545	Ile	Ala	His	Pro	Val 550	Ser	Asp	Asp	Ser	Asp 555	Leu	Pro	Glu	Ser	Ala 560	
15		Ser	Ser	Pro	Ala	Ala 565	Gly	Pro	Thr	Lys	Pro 570	Ala	Ala	Ser	Gln	Leu 575	Glu	
		Ser	Asp	Thr	Ile 580	Ala	Asp	Leu	Pro	Asp 585		Thr	Val	Val	Thr 590		Ser	
		Thr	Asn	Asp 595		His	Asp	Val	Val 600	Val	Val	Asp	Val	Glu 605	Asp	Asp	Pro	
20		Asp	Glu 610		Ala	Val												
	(218) IN										•						
25		(1	(B) LE) TY) ST	NGTH PE :	: 18 nucl EDNE	54 b eic SS:	ase acid sing	pair	s								
		•	.i) M					_										
	;	(x	i) S	EQUE	NCE	DESC	RIPT	'ION:	SEQ	ID	NO : 2	17:						
30	ATGG	GGCC	CA C	CCTA	.GCGG	T TC	CCAC	cccc	TAT	GGCT	GTA	TTGG	CTGT	AA G	CTAC	CCCA	G	60
	CCAG	ATA	CC C	ACCG	GCTC	T AA	TCAT	CTTT	ATG	TTCT	'GCG	CGAT	GGTT	AT C	ACCA	TCGT	T	120
	GTAG	BACCI	r aa:	CGGC	AACT	C CA	TGGT	CATT	TTG	GCTG	TGA	CGAA	GAAC	AA G	AAGC	TCCG	G	180
	AATT	CTG	CA A	CATC	TTCG	T GG	TCAG	TCTC	TCT	GTGG	CCG	TATA	GCTG	GT G	GCCA	TCTA	.C	240
	CCAT	TACCO	TT 1	GATG	CTGC	A TG	CCAT	GTCC	TTA:	'GGGG	GCT	GGGA	TCTG	AG C	CAGI	TACA	.G	300
35	TGCC	AGAT	rgg 1	'CGGG	TTCA	T CA	CAGG	GCTG	AGI	GTGG	TCG	GCTC	CATO	TT C	AACA	TCGT	G	360

	GCAATCGCTA	TCAACCGTTA	CTGCTACATC	TGCCACAGCC	TCCAGTACGA	ACGGATCTTC	420
	AGTGTGCGCA	ATACCTGCAT	CTACCTGGTC	ATCACCTGGA	TCATGACCGT	CCTGGCTGTC	480
	CTGCCCAACA	TGTACATTGG	CACCATCGAG	TACGATCCTC	GCACCTACAC	CTGCATCTTC	540
	AACTATCTGA	ACAACCCTGT	CTTCACTGTT	ACCATCGTCT	GCATCCACTT	CGTCCTCCCT	600
5	CTCCTCATCG	TGGGTTTCTG	CTACGTGAGG	ATCTGGACCA	AAGTGCTGGC	GGCCCGTGAC	660
	CCTGCAGGGC	AGAATCCTGA	CAACCAACTT	GCTGAGGTTC	GCAATAAACT	AACCATGTTT	720
	GTGATCTTCC	TCCTCTTTGC	AGTGTGCTGG	TGCCCTATCA	ACGTGCTCAC	TGTCTTGGTG	780
	GCTGTCAGTC	CGAAGGAGAT	GGCAGGCAAG	ATCCCCAACT	GGCTTTATCT	TGCAGCCTAC	840
	TTCATAGCCT	ACTTCAACAG	CTGCCTCAAC	GCTGTGATCT	ACGGGCTCCT	CAATGAGAAT	900
10	TTCCGAAGAG	AATACTGGAC	CATCTTCCAT	GCTATGCGGC	ACCCTATCAT	ATTCTTCTCT	960
	GGCCTCATCA	GTGATATTCG	TGAGATGCAG	GAGGCCCGTA	CCCTGGCCCG	CGCCCGTGCC	1020
	CATGCTCGCG	ACCAAGCTCG	TGAACAAGAC	CGTGCCCATG	CCTGTCCTGC	TGTGGAGGAA	1080
	ACCCCGATGA	ATGTCCGGAA	TGTTCCATTA	CCTGGTGATG	CTGCAGCTGG	CCACCCGAC	1140
	CGTGCCTCTG	GCCACCCTAA	GCCCCATTCC	AGATCCTCCT	CTGCCTATCG	CAAATCTGCC	1200
15	TCTACCCACC	ACAAGTCTGT	CTTTAGCCAC	TCCAAGGCTG	CCTCTGGTCA	CCTCAAGCCT	1260
	GTCTCTGGCC	ACTCCAAGCC	TGCCTCTGGT	CACCCCAAGT	CTGCCACTGT	CTACCCTAAG	1320
	CCTGCCTCTG	TCCATTTCAA	GGCTGACTCT	GTCCATTTCA	AGGGTGACTC	TGTCCATTTC	1380
	AAGCCTGACT	CTGTTCATTT	CAAGCCTGCT	TCCAGCAACC	CCAAGCCCAT	CACTGGCCAC	1440
	CATGTCTCTG	CTGGCAGCCA	CTCCAAGTCT	GCCTTCAGTG	CTGCCACCAG	CCACCCTAAA	1500
20	CCCACCACTG	GCCACATCAA	GCCAGCTACC	AGCCATGCTG	AGCCCACCAC	TGCTGACTAT	1560
	CCCAAGCCTG	CCACTACCAG	CCACCCTAAG	CCCACTGCTG	CTGACAACCC	TGAGCTCTCT	1620
	GCCTCCCATT	GCCCCGAGAT	CCCTGCCATT	GCCCACCCTG	TGTCTGACGA	CAGTGACCTC	1680
	CCTGAGTCGG	CCTCTAGCCC	TGCCGCTGGG	CCCACCAAGC	CTGCTGCCAG	CCAGCTGGAG	1740
	TCTGACACCA	TCGCTGACCT	TCCTGACCCT	ACTGTAGTCA	CTACCAGTAC	CAATGATTAC	1800
25	CATGATGTCG	TGGTTGTTGA	TGTTGAAGAT	GATCCTGATG	AAATGGCTGT	GTGA	1854
	()						

(219) INFORMATION FOR SEQ ID NO:218:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 617 amino acids
 - (B) TYPE: amino acid

- (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:218:

٠.		•		, .												
5	Met 1	Gly	Pro	Thr	Leu 5	Ala	Val	Pro	Thr	Pro 10	Tyr	Gly	Cys	Ile	Gly 15	Cys
	Lys	Leu	Pro	Gln 20	Pro	Glu	Tyr	Pro	Pro 25	Ala	Leu	Ile	Ile	Phe 30	Met	Phe
10	Cys	Ala	Met 35	Val	Ile	Thr	Ile	Val 40	Val	Asp	Leu	Ile	Gly 45	Asn	Ser	Met
	Val	Ile 50	Leu	Ala	Val	Thr	Lys 55	Asn	Lys	Lys	Leu	Arg 60	Asn	Ser	Gly	Asn
	Ile 65	Phe	Val	Val	Ser	Leu 70	Ser	Val	Ala	Asp	Met 75	Leu	Val	Ala	Ile	Tyr 80
15	Pro	Tyr	Pro	Leu	Met 85	Leu	His	Ala	Met	Ser 90	Ile	Gly	Gly	Trp	Asp 95	Leu
	Ser	Gln	Leu	Gln 100	Cys	Gln	Met	Val	Gly 105	Phe	Ile	Thr	Gly	Leu 110	Ser	Val
20	Val	Gly	Ser 115	Ile	Phe	Asn	Ile	Val 120	Ala	Ile	Ala	Ile	Asn 125	Arg	Tyr	Cys
	Tyr	Ile 130	Cys	His	Ser	Leu	Gln 135	Tyr	Glu	Arg	Ile	Phe 140	Ser	Val	Arg	Asn
	Thr 145	Cys	Ile	Tyr	Leu	Val 150	Ile	Thr	Trp	Ile	Met 155	Thr	Val	Leu	Ala	Val 160
25	Leu	Pro	Asn	Met	Tyr 165	Ile	Gly	Thr	Ile	Glu 170	Tyr	Asp	Pro	Arg	Thr 175	Tyr
	Thr	Cys	Ile	Phe 180	Asn	Tyr	Leu	Asn	Asn 185	Pro	Val	Phe	Thr	Val 190	Thr	Ile
30	Val	Cys	Ile 195	His	Phe	Val	Leu	Pro 200	Leu	Leu	Ile	Val	Gly 205	Phe	Cys	Tyr
	Val	Arg 210	Ile	Trp	Thr	Lys	Val 215	Leu	Ala	Ala	Arg	Asp 220	Pro	Ala	Gly	Gln
	Asn 225	Pro	Asp	Asn	Gln	Leu 230	Ala	Glu	Val	Arg	Asn 235	Lys	Leu	Thr		Phe 240
35	Val	Ile	Phe	Leu	Leu 245	Phe	Ala	Val	Cys	Trp 250	Cys	Pro	Ile	Asn	Val 255	Leu

	Thr	Val	Leu	Val 260	Ala	Val	Ser	Pro	Lys 265	Glu	Met	Ala	Gly	Lys 270	Ile	Pro
	Asn	Trp	Leu 275	Tyr	Leu	Ala	Ala	Tyr 280	Phe	Ile	Ala	Tyr	Phe 285	Asn	Ser	Cys
5	Leu	Asn 290	Ala	Val	Ile	Tyr	Gly 295	Leu	Leu	Asn	Glu	Asn 300		Arg	Arg	Glu
	Tyr 305	Trp	Thr	Ile	Phe	His 310	Ala	Met	Arg	His	Pro 315	Ile	Ile	Phe	Phe	Ser 320
10	Gly	Leu	Ile	Ser	Asp 325	Ile	Arg	Glu	Met	Gln 330	Glu	Ala	Arg	Thr	Leu 335	Ala
	Arg	Ala	Arg	Ala 340		Ala	Arg	Asp	Gln 345	Ala	Arg	Glu	Gln	Asp 350	Arg	Ala
	His	Ala	Cys 355	Pro	Ala	Val	Glu	Glu 360	Thr	Pro	Met	Asn	Val 365	Arg	Asn	Val
15	Pro	Leu 370	Pro	Gly	Asp	Ala	Ala 375	Ala	Gly	His	Pro	Asp 380	Arg	Ala	Ser	Gly
	His 385	Pro	Lys	Pro	His	Ser 390	Arg	Ser	Ser	Ser	Ala 395	Tyr	Arg	Lys	Ser	Ala 400
20	Ser	Thr	His	His	Lys 405	Ser	Val	Phe	Ser	His 410	Ser	Lys	Ala		Ser 415	Gly
	His	Leu	Lys	Pro 420	Val	Ser	Gly	His	Ser 425	Lys	Pro	Ala	Ser	Gly 430	His	Pro
	Lys	Ser	Ala 435	Thr	Val	Tyr	Pro	Lys 440	Pro	Ala	Ser	Val	His 445	Phe	Lys	Ala
25	Asp	Ser 450	Val	His	Phe	Lys	Gly 455	Asp	Ser	Val	His	Phe 460	Lys	Pro	Asp	Ser
	Val 465		Phe	Lys	Pro	Ala 470	Ser	Ser	Asn	Pro	Lys 475	Pro	Ile	Thr	Gly	His 480
30	His	Val	Ser	Ala	Gly 485	Ser	His	Ser	Lys	Ser 490	Ala	Phe	Ser	Ala	Ala 495	Thr
٠.,	Ser	His	Pro	Lys 500	Pro	Thr	Thr	Gly	His 505	Ile	Lys	Pro	Ala	Thr 510	Ser	His
	Ala	Glu	Pro 515	Thr	Thr	Ala	Asp	Tyr 520	Pro	Lys	Pro	Ala	Thr 525	Thr	Ser	His
35	Pro	Lys 530	Pro	Thr	Ala	Ala	Asp 535	Asn	Pro	Glu	Leu	Ser 540	Ala	Ser	His	Cys
	Pro	Glu	Ile	Pro	Ala	Ile	Ala	His	Pro	Val	Ser	Asp	Asp	Ser	Asp	Leu

	545		550	. 1	55	55		560
	Pro G	lu Ser Ala	Ser Ser P 565	ro Ala Al	a Gly Pi 570	o Thr Lys	Pro Ala 575	Ala
5	Ser G	ln Leu Glu 580	Ser Asp T			eu Pro As <u>r</u>		Val
	Val T	hr Thr Ser 595	Thr Asn A	sp Tyr Hi 600	s Asp Va	ıl Val Val 605		Val
		sp Asp Pro 10	Asp Glu M		1		ſ	
10	(220) INFO	RMATION FOR	SEQ ID NO	0:219:				
15	2.134 (2.11)	SEQUENCE CH (A) LENGTH: (B) TYPE: n (C) STRANDE (D) TOPOLOG	1548 base ucleic ac: DNESS: si	e pairs id ngle		i de la companya de l		s.
	(ii)	MOLECULE T	YPE: DNA	(genomic)				
	(xi)	SEQUENCE D	ESCRIPTION	N: SEQ ID	NO:219:			
	ATGGGACATA	ACGGGAGCTG	GATCTCTCC	CA AATGCC	AGCG AGC	CGCACAA C	GCGTCCGGC	60
	GCCGAGGCTG	CGGGTGTGAA	CCGCAGCG	CG CTCGGG	GAGT TCG	GCGAGGC G	CAGCTGTAC	120
20	CGCCAGTTCA	CCACCACCGT	GCAGGTCGT	C ATCTTC	ATAG GCT	CGCTGCT C	GGAAACTTC	180
	ATGGTGTTAT	GGTCAACTTG	CCGCACAA	CC GTGTTC	AAAT CTG	TCACCAA C	AGGTTCATI	240
	AAAAACCTGG	CCTGCTCGGG	GATTTGTG	CC AGCCTG	STCT GTG	TGCCCTT C	GACATCATC	300
	CTCAGCACCA	GTCCTCACTG	TTGCTGGT	G ATCTAC	ACCA TGC	TCTTCTG C	AAGGTCGTC	360
	AAATTTTTGC	ACAAAGTATT	CTGCTCTGT	G ACCATC	CTCA GCT	TCCCTGC T	ATTGCTTTG	420
25	GACAGGTACT	ACTCAGTCCT	CTATCCACT	G GAGAGG	TAT AAAA	CTGATGC C	AAGTCCCGT	480
	GAACTGGTGA	TGTACATCTG	GGCCCATGC	CA GTGGTG	GCCA GTG	TCCCTGT G	TTTGCAGTA	540
	ACCAATGTGG	CTGACATCTA	TGCCACGTC	C ACCTGC	ACGG AAG	TCTGGAG C	AACTCCTTG	600
	GGCCACCTGG	TGTACGTTCT	GGTGTATA	C ATCACC	ACGG TCA	TTGTGCC T	GTGGTGGTG	660
	GTGTTCCTCT	TCTTGATACT	GATCCGACG	G GCCTG	AGTG CCA	GCCAGAA G	AAGAAGGTC	720
30	ATCATAGCAG	CGCTCCGGAC	CCCACAGAA	C ACCATC	CTA TTC	CCTATGC C	TCCCAGCGG	780
	GAGGCCGAGC	TGAAAGCCAC	CCTGCTCTC	C ATGGTG	ATGG TCT	TCATCTT G	TGTAGCGTG	840
	CCCTATGCCA	CCCTGGTCGT	CTACCAGAC	T GTGCTC	AATG TCC	CTGACAC T	TCCGTCTTC	900

WO 00/22129 PCT/US99/23938

	TTGCTGCTCA CTGCTGTTTG GCTGCCCAAA GTCTCCCTGC TGGCAAACCC TGTTCTCTTT	960
•	CTTACTGTGA ACAAATCTGT CCGCAAGTGC TTGATAGGGA CCCTGGTGCA ACTACACCAC	1020
	CGGTACAGTC GCCGTAATGT GGTCAGTACA GGGAGTGGCA TGGCTGAGGC CAGCCTGGAA	1080
÷	CCCAGCATAC GCTCGGGTAG CCAGCTCCTG GAGATGTTCC ACATTGGGCA GCAGCAGATC	1140
5	TTTAAGCCCA CAGAGGATGA GGAAGAGAGT GAGGCCAAGT ACATTGGCTC AGCTGACTTC	1200
	CAGGCCAAGG AGATATTTAG CACCTGCCTG GAGGGAGAGC AGGGGCCACA GTTTGCGCCC	1260
	TCTGCCCCAC CCCTGAGCAC AGTGGACTCT GTATCCCAGG TGGCACCGGC AGCCCCTGTG	1320
	GAACCTGAAA CATTCCCTGA TAAGTATTCC CTGCAGTTTG GCTTTGGGCC TTTTGAGTTG	1380
	CCTCCTCAGT GGCTCTCAGA GACCCGAAAC AGCAAGAAGC GGCTGCTTCC CCCCTTGGGC	1440
10	AACACCCCAG AAGAGCTGAT CCAGACAAAG GTGCCCAAGG TAGGCAGGGT GGAGCGGAAG	1500
	ATGAGCAGAA ACAATAAAGT GAGCATTTTT CCAAAGGTGG ATTCCTAG	1548
	(221) INFORMATION FOR SEQ ID NO:220:	••
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 515 amino acids	
15	(B) TYPE: amino acid (C) STRANDEDNESS:	•
	(D) TOPOLOGY: not relevant	•
	(ii) MOLECULE TYPE: protein	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:220:	
20	Met Gly His Asn Gly Ser Trp Ile Ser Pro Asn Ala Ser Glu Pro Hi 1 5 10 15	is
	Asn Ala Ser Gly Ala Glu Ala Ala Gly Val Asn Arg Ser Ala Leu G 20 25 30	ly
25	Glu Phe Gly Glu Ala Gln Leu Tyr Arg Gln Phe Thr Thr Val Gl	ln
	Val Val Ile Phe Ile Gly Ser Leu Leu Gly Asn Phe Met Val Leu Tr 50 55 60	сp
	Ser Thr Cys Arg Thr Thr Val Phe Lys Ser Val Thr Asn Arg Phe II	
30	Lys Asn Leu Ala Cys Ser Gly Ile Cys Ala Ser Leu Val Cys Val Pr 85 90 95	ro
	Phe Asp Ile Ile Leu Ser Thr Ser Pro His Cys Cys Trp Trp Ile Ty	yr

•	Thr	Met	Leu 115	Phe	Cys	Lys	Val	Val 120	Lys	Phe	Leu	His	Lys 125	Val	Phe	Cys
	Ser	Val 130	Thr	Ile	Leu	Ser	Phe 135	Pro	Ala	Ile	Ala	Leu 140	Asp	Arg	Tyr	Tyr
5	Ser 145	Val	Leu	Tyr	Pro	Leu 150	Glu	Arg	Lys	Ile	Ser 155	Asp	Ala	Lys	Ser	Arg 160
	Glu	Leu	Val	Met	Tyr 165	Ile	Trp	Ala	His	Ala 170	Val	Val	Ala	Ser	Val 175	Pro
10	Val	Phe	Ala	Val 180	Thr	Asn	Val	Ala	Asp 185	Ile	Tyr	Ala	Thr	Ser 190	Thr	Cys
	Thr	Glu	Val 195	Trp	Ser	Asn	Ser	Leu 200	Gly	His	Leu	Val	Tyr 205	Val	Leu	Val
	Tyr	Asn 210	Ile	Thr	Thr	Val	Ile 215	Val	Pro	Val	Val	Val 220	Val	Phe	Leu	Phe
15	Leu 225	Ile	Leu	Ile	Arg	Arg 230	Ala	Leu	Ser	Ala	Ser 235	Gln	Lys	Lys	Lys	Val 240
		Ile			245					250			·		255	
20		Ser		260			••		265					270		
		Val	275			-		280					285			
		290					295					300				Thr
25	305					310					315		*	•		Phe 320
		Thr			325					330					335	
30		Leu		340					345					350		
	_		355					360		•			365	•		Gln
35		370					375					380				Phe
<i>33</i>	385					390					395					400 Pro

	• • • •			•	405	* * *				410	•			•	415		٠
	Gln	Phe	Ala	Pro 420	Ser	Ala	Pro	Pro	Leu 425	Ser	Thr	Val	Asp	Ser 430	Val	Ser	
5	Gln	Val	Ala 435	Pro	Ala	Ala	Pro	Val 440	Glu	Pro	Glu	Thr	Phe 445		Asp	Lys	
	Tyr	Ser 450	Leu	Gln	Phe	Gly	Phe 455	Gly	Pro	Phe	Glu	Leu 460	Pro	Pro	Gln	Trp	
	Leu 465	Ser	Glu	Thr	Arg	Asn 470	Ser	Lys	Lys	Arg	Leu 475	Leu	Pro	Pro	Leu	Gly 480	
10	Asn	Thr	Pro	Glu	Glu 485	Leu	Ile	Gln	Thr	Lys 490	Val	Pro	Lys	Val	Gly 495	Arg	
	Val	Glu	Arg	Lys 500	Met	Ser	Arg	Asn	Asn 505		Val	Ser	Ile	Phe 510	Pro	Lys	
15	Val	Asp	Ser 515														
	(222) IN	FORM	ATIO	N FO	R SE	Q ID	NO:	221:									
20	(i	(A (B (C (D	QUEN LE TY ST TO	NGTH PE: : RAND POLO	: 11 nucl EDNE GY: TYPE	64 beeic assistant series of the series of t	ase pacid sing ar A (g	pair: le enom	ic)							,	
			EQUE													_	
	ATGAATCG	•															60
25	TTCCGAGA																120
	GGGCTTCT																180
	TCCAGCCG	GA T	TTTC	CTGT	T CA	ACCT	GGCA	GTA	GCTG	ACT	TTCT	ACTG	AT C	ATCT	GCCT	G	240
	CCGTTCGT	GA I	GGAC	TACT	A TG	TGCG	GCGT	TCA	GACT	GGA	AGTT	TGGG	GA C	ATCC	CTTG	С	300
	CGGCTGGT	GC I	CTTC	ATGT	T TG	CCAT	GAAC	CGC	CAGG	GCA	GCAI	CATC	TT C	CTCA	CGGT	G	360
30	GTGGCGGT	AG A	CAGG	TATT	T CC	:GGGT	GGTC	CAT	cccc	ACC	ACGC	CCTG	AA C	AAGA	TCTC	C	420
	AATTGGAC	CAG C	CAGCO	'ATCA	T CI	CTTG	CCTI	CTG	TGGG	GCA	TCAC	TGTT	GG C	CTAA	.CAGT	C	480
	CACCTCCT	rga A	AGAAG	AAGI	T GO	TGAT	CCAG	TAA	GGCC	CTG	CAAA	TGTG	TG C	ATCA	GCTT.	'C	540
	AGCATCTO	בכר ז	ነግል ሮ ር	ידידיכיכ	G GI	GGCA	CGAA	GCI	ATGI	TCC	TCCI	GGAG	TT C	CTCC	TGCC	:C	600

	CIGGGCAICA ICCIGITCIG CTCAGCCAGA ATTATCTGGA GCCTGCGGCA GAGACAAATG 66
	GACCGGCATG CCAAGATCAA GAGAGCCAAA ACCTTCATCA TGGTGGTGGC CATCGTCTTT 72
	GTCATCTGCT TCCTTCCCAG CGTGGTTGTG CGGATCCGCA TCTTCTGGCT CCTGCACACT 78
	TCGGGCACGC AGAATTGTGA AGTGTACCGC TCGGTGGACC TGGCGTTCTT TATCACTCTC 84
. 5	AGCTTCACCT ACATGAACAG CATGCTGGAC CCCGTGGTGT ACTACTTCTC CAGCCCATCC 900
	TTTCCCAACT TCTTCTCCAC TTTGATCAAC CGCTGCCTCC AGAGGAAGAT GACAGGTGAG 960
	CCAGATAATA ACCGCAGCAC GAGCGTCGAG CTCACAGGGG ACCCCAACAA AACCAGAGGC 1020
	GCTCCAGAGG CGTTAATGGC CAACTCCGGT GAGCCATGGA GCCCCTCTTA TCTGGGCCCA 1080
:	ACCTCAAATA ACCATTCCAA GAAGGGACAT TGTCACCAAG AACCAGCATC TCTGGAGAAA 1140
10	CAGTTGGGCT GTTGCATCGA GTAA 1164
	(223) INFORMATION FOR SEQ ID NO:222:
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 387 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: (D) TOPOLOGY: not relevant
	(ii) MOLECULE TYPE: protein
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:222:
20	Met Asn Arg His His Leu Gln Asp His Phe Leu Glu Ile Asp Lys Lys 1 5 10 15
	Asn Cys Cys Val Phe Arg Asp Asp Phe Ile Ala Lys Val Leu Pro Pro 20 25 30
	Val Leu Gly Leu Glu Phe Ile Phe Gly Leu Leu Gly Asn Gly Leu Ala 35 40 45
25	Leu Trp Ile Phe Cys Phe His Leu Lys Ser Trp Lys Ser Ser Arg Ile 50 55 60
	Phe Leu Phe Asn Leu Ala Val Ala Asp Phe Leu Leu Ile Ile Cys Leu 65 70 75 80
30	Pro Phe Val Met Asp Tyr Tyr Val Arg Arg Ser Asp Trp Lys Phe Gly 85 90 95
	Asp Ile Pro Cys Arg Leu Val Leu Phe Met Phe Ala Met Asn Arg Gln 100 105 110
	Gly Ser Ile Ile Phe Leu Thr Val Val Ala Val Asp Arg Tyr Phe Arg 115 120 125

		Val	Val 130	His	Pro	His	His	Ala 135	Leu	Asn	Lys	Ile	Ser 140	Asn	Trp	Thr	Ala
		Ala 145	Ile	Ile	Ser	Cys	Leu 150	Leu	Trp	Gly	Ile	Thr 155	Val	Gly	Leu	Thr	Val 160
5		His	Leu	Leu	Lys	Lys 165	Lys	Leu	Leu	Ile	Gln 170	Asn	Gly	Pro	Ala	Asn 175	Val
		Cys	Ile	Ser	Phe 180	Ser	Ile	Cys	His	Thr 185	Phe	Arg	Trp	His	Glu 190	Ala	Met
10		Phe	Leu	Leu 195	Glu	Phe	Leu	Leu	Pro 200	Leu	Gly	Ile	Ile	Leu 205	Phe	Cys	Ser
	,	Ala	Arg 210	Ile	Ile	Trp	Ser	Leu 215	Arg	Gln	Arg	Gln	Met 220	Asp	Arg	His	Ala
		Lys 225	Ile	Lys	Arg	Ala	Lys 230	Thr	Phe	Ile	Met	Val 235	Val	Ala	Ile	Val	Phe 240
15	*	Val	Ile	Cys	Phe	Leu 245	Pro	Ser	Val	Val	Val 250	Arg	Ile	Arg	Ile	Phe 255	Trp
		Leu	Leu	His	Thr 260	Ser	Gly	Thr	Gln	Asn 265	Cys	Glu	Val	Tyr	Arg 270	Ser	Val
20		Asp	Leu	Ala 275	Phe	Phe	Ile	Thr	Leu 280	Ser	Phe	Thr	Tyr	Met 285	Asn	Ser	Met
		Leu	Asp 290	Pro	Val	Val	Tyr	Tyr 295	Phe	Ser	Ser	Pro	Ser 300	Phe	Pro	Asn	Phe
	·	Phe 305	Ser	Thr	Leu	Ile	Asn 310	Arg	Cys	Leu	Gln	Arg 315	Lys	Met	Thr	Gly	Glu 320
25		Pro	Asp	Asn	Asn	Arg 325	Ser	Thr	Ser	Val	Glu 330	Leu	Thr	Gly	Asp	Pro 335	Asn
		Lys	Thr	Arg	Gly 340	Ala	Pro	Glu	Ala	Leu 345	Met	Ala	Asn	Ser	Gly 350	Glu	Pro
30		Trp	Ser	Pro 355	Ser	Tyr	Leu	Gly	Pro 360	Thr	Ser	Asn	Asn	His 365	Ser	Lys	Lys
		Gly	His 370	Cys	His	Gln	Glu	Pro 375	Ala	Ser	Leu	Glu	Lys 380	Gln	Leu	Gly	Cys
		Cys 385	Ile	Glu										·			
35	(224)	INI	FORM	OITA	I FOI	R SEC) İD	NO: 2	223:								

- 35 (224) INFORMATION FOR SEQ ID NO:223:
 - (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1212 base pairs

(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic) 5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:223: ATGGCTTGCA ATGGCAGTGC GGCCAGGGGG CACTTTGACC CTGAGGACTT GAACCTGACT 60 GACGAGGCAC TGAGACTCAA GTACCTGGGG CCCCAGCAGA CAGAGCTGTT CATGCCCATC TGTGCCACAT ACCTGCTGAT CTTCGTGGTG GGCGCTGTGG GCAATGGGCT GACCTGTCTG 180 GTCATCCTGC GCCACAAGGC CATGCGCACG CCTACCAACT ACTACCTCTT CAGCCTGGCC GTGTCGGACC TGCTGGTGCT GCTGGTGGGC CTGCCCCTGG AGCTCTATGA GATGTGGCAC 300 AACTACCCCT TCCTGCTGGG CGTTGGTGGC TGCTATTTCC GCACGCTACT GTTTGAGATG GTCTGCCTGG CCTCAGTGCT CAACGTCACT GCCCTGAGCG TGGAACGCTA TGTGGCCGTG 420 GTGCACCCAC TCCAGGCCAG GTCCATGGTG ACGCGGGCCC ATGTGCGCCG AGTGCTTGGG 480 GCCGTCTGGG GTCTTGCCAT GCTCTGCTCC CTGCCCAACA CCAGCCTGCA CGGCATCCGG 540 15 CAGCTGCACG TGCCCTGCCG GGGCCCAGTG CCAGACTCAG CTGTTTGCAT GCTGGTCCGC CCACGGGCCC TCTACAACAT GGTAGTGCAG ACCACCGCGC TGCTCTTCTT CTGCCTGCCC ATGGCCATCA TGAGCGTGCT CTACCTGCTC ATTGGGCTGC GACTGCGGCG GGAGAGGCTG 720 CTGCTCATGC AGGAGGCCAA GGGCAGGGGC TCTGCAGCAG CCAGGTCCAG ATACACCTGC 780 AGGCTCCAGC AGCACGATCG GGGCCGGAGA CAAGTGAAGA AGATGCTGTT TGTCCTGGTC 840 20 GTGGTGTTTG GCATCTGCTG GGCCCCGTTC CACGCCGACC GCGTCATGTG GAGCGTCGTG 900 TCACAGTGGA CAGATGGCCT GCACCTGGCC TTCCAGCACG TGCACGTCAT CTCCGGCATC 960 TTCTTCTACC TGGGCTCGGC GGCCAACCCC GTGCTCTATA GCCTCATGTC CAGCCGCTTC 1020 CGAGAGACCT TCCAGGAGGC CCTGTGCCTC GGGGCCTGCT GCCATCGCCT CAGACCCCGC 1080 CACAGCTCCC ACAGCCTCAG CAGGATGACC ACAGGCAGCA CCCTGTGTGA TGTGGGCTCC 1140 25 CTGGGCAGCT GGGTCCACCC CCTGGCTGGG AACGATGGCC CAGAGGCGCA GCAAGAGACC 1200 GATCCATCCT GA 1212

(225) INFORMATION FOR SEQ ID NO:224:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 403 amino acids

30 (B) TYPE: amino acid

ב באו האפרכנית חשב חוייחופתה

- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:224:
- Met Ala Cys Asn Gly Ser Ala Ala Arg Gly His Phe Asp Pro Glu Asp 5 Leu Asn Leu Thr Asp Glu Ala Leu Arg Leu Lys Tyr Leu Gly Pro Gln 25 Gln Thr Glu Leu Phe Met Pro Ile Cys Ala Thr Tyr Leu Leu Ile Phe 10 40 Val Val Gly Ala Val Gly Asn Gly Leu Thr Cys Leu Val Ile Leu Arg His Lys Ala Met Arg Thr Pro Thr Asn Tyr Tyr Leu Phe Ser Leu Ala Val Ser Asp Leu Leu Val Leu Leu Val Gly Leu Pro Leu Glu Leu Tyr 15 90 85 Glu Met Trp His Asn Tyr Pro Phe Leu Leu Gly Val Gly Gly Cys Tyr 105 Phe Arg Thr Leu Leu Phe Glu Met Val Cys Leu Ala Ser Val Leu Asn 20 Val Thr Ala Leu Ser Val Glu Arg Tyr Val Ala Val His Pro Leu 135 130 Gln Ala Arg Ser Met Val Thr Arg Ala His Val Arg Arg Val Leu Gly 155 150 Ala Val Trp Gly Leu Ala Met Leu Cys Ser Leu Pro Asn Thr Ser Leu 25 165 His Gly Ile Arg Gln Leu His Val Pro Cys Arg Gly Pro Val Pro Asp 185 Ser Ala Val Cys Met Leu Val Arg Pro Arg Ala Leu Tyr Asn Met Val . 30 200 Val Gln Thr Thr Ala Leu Leu Phe Phe Cys Leu Pro Met Ala Ile Met 215 210 Ser Val Leu Tyr Leu Leu Ile Gly Leu Arg Leu Arg Arg Glu Arg Leu 235

Leu Leu Met Gln Glu Ala Lys Gly Arg Gly Ser Ala Ala Ala Arg Ser

	Arg	Tyr	Thr	Cys 260	Arg	Leu	Gln	Gln	His 265	Asp	Arg	Gly	Arg	Arg 270	Gln	Val
	Lys	Lys	Met 275	Leu	Phe	Val	Leu	Val 280	Val	Val	Phe	Gly	Ile 285	Cys	Trp	Ala
5	Pro	Phe 290	His	Ala	Asp	Arg	Val 295	Met	Trp	Ser	Val	Val	Ser	Gln	Trp	Thr
	Asp 305	Gly	Leu	His	Leu	Ala 310	Phe	Gln	His	Val	His 315	Val	Ile	Ser	Gly	Ile 320
10	Phe	Phe	Tyr	Leu	Gly 325	Ser	Ala	Ala	Asn	Pro 330	Val	Leu	Tyr	Ser	Leu 335	Met
	Ser	Ser	Arg	Phe 340	Arg	Glu	Thr	Phe	Gln 345	Glu	Ala	Leu	Cys	Leu 350	Gly	Ala
	Cys	Cys	His 355	Arg	Leu	Arg	Pro	Arg 360	His	Ser	Ser	His	Ser 365	Leu	Ser	Arg
15	Met	Thr 370	Thr	Gly	Ser	Thr	Leu 375	Cys	Asp	Val	Gly	Ser. 380	Leu	Gly	Ser	Trp
	Val 385	His	Pro	Leu	Ala	Gly 390	Asn	Asp	Gly	Pro	Glu 395	Ala	Gln	Gln		Thr 400
20	Asp	Pro	Ser			•					,					
	(226) IN	FORMA	TION	FOR	SEQ	ID	NO:2	25:								
25	(i)	(B) (C)		GTH: E: n ANDE	109 ucle DNES	8 ba ic a S: s	se p cid ingl	airs		;						
	(ii) MO	LECU	LE T	YPE:	DNA	(ge	nomi	c)							
•	(xi	.) SE	QUEN	CE D	ESCR	IPTI	ON:	SEQ	ID N	0:22	5:					
	ATGGGGAAC	A TC	ACTG	CAGA	CAA	CTCC	TCG .	ATGA	GCTG'	TA C	CATC	GACC	A TA	CCAT	CCAC	60
30	CAGACGCTG	G CC	CCGG	TGGT	CTA	TGTT	ACC (GTGC'	TGGT	GG T	GGGC'	rtcc	C GG	CCAA	CTGC	120
	CTGTCCCTC	T AC	TTCG	GCTA	CCT	GCAG:	ATC :	AAGG	CCCG	GA A	CGAG	CTGG	G CG	TGTA	CCTG	180
	TGCAACCTG	A CG	GTGG	CCGA	CCT	CTTC'	TAC I	ATCT	GCTC	GC T	GCCC'	rtct	G GC	TGCA(GTAC	240
	GTGCTGCAG	C AC	GACA	ACTG	GTC'	rcac(GGC (GACC'	TGTC	CT G	CCAG	STGT	G CG	GCAT	CCTC	300
	CTGTACGAG															
35	CTGGCTGTG	G CC	CATC	CCTT	CCG	CTTC	CAC (CAGT:	rccgo	SA CO	CTG	AAGGO	GGG	CCGT	CGC	420

ENZUCCIU -MU UUSSISGII IV-

PCT/US99/23938

15

30

191

	GTCAGCGTGG	TCATCTGGGC	CAAGGAGCTG	CTGACCAGCA	TCTACTTCCT	GATGCACGAG	480
	GAGGTCATCG	AGGACGAGAA	CCAGCACCGC	GTGTGCTTTG	AGCACTACCC	CATCCAGGCA	540
	TGGCAGCGCG	CCATCAACTA	CTACCGCTTC	CTGGTGGGCT	TCCTCTTCCC	CATCTGCCTG	600
	CTGCTGGCGT	CCTACCAGGG	CATCCTGCGC	GCCGTGCGCC	GGAGCCACGG	CACCCAGAAG	660
5	AGCCGCAAGG	ACCAGATCAA	GCGGCTGGTG	CTCAGCACCG	TGGTCATCTT	CCTGGCCTGC	720
	TTCCTGCCCT	ACCACGTGTT	GCTGCTGGTG	CGCAGCGTCT	GGGAGGCCAG	CTGCGACTTC	780
	GCCAAGGGCG	TTTTCAACGC	CTACCACTTC	TCCCTCCTGC	TCACCAGCTT	CAACTGCGTC	840
	GCCGACCCCG	TGCTCTACTG	CTTCGTCAGC	GAGACCACCC	ACCGGGACCT	GGCCCGCCTC	900
	CGCGGGGCCT	GCCTGGCCTT	CCTCACCTGC	TCCAGGACCG	GCCGGGCCAG	GGAGGCCTAC	960
10	CCGCTGGGTG	CCCCGAGGC	CTCCGGGAAA	AGCGGGGCCC	AGGGTGAGGA	GCCCGAGCTG	1020
	TTGACCAAGC	TCCACCCGGC	CTTCCAGACC	CCTAACTCGC	CAGGGTCGGG	CGGGTTCCCC	1080
,	ACGGGCAGGT	TGGCCTAG		·			1098
	()		CEO ID NO.	226.			

(227) INFORMATION FOR SEQ ID NO:226:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 365 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein
- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:226:

Met Gly Asn Ile Thr Ala Asp Asn Ser Ser Met Ser Cys Thr Ile Asp 1 5 10 15

His Thr Ile His Gln Thr Leu Ala Pro Val Val Tyr Val Thr Val Leu 20 25 30

25 Val Val Gly Phe Pro Ala Asn Cys Leu Ser Leu Tyr Phe Gly Tyr Leu 35 40 45

Gln Ile Lys Ala Arg Asn Glu Leu Gly Val Tyr Leu Cys Asn Leu Thr 50 55 60

Val Ala Asp Leu Phe Tyr Ile Cys Ser Leu Pro Phe Trp Leu Gln Tyr 65 70 75 80

Val Leu Gln His Asp Asn Trp Ser His Gly Asp Leu Ser Cys Gln Val 85 90 95

Cys Gly Ile Leu Leu Tyr Glu Asn Ile Tyr Ile Ser Val Gly Phe Leu

				100					105					110		
	Cys	Cys	Ile 115	Ser	Val	Asp	Arg	Tyr 120	Leu	Ala	Val	Ala	His 125	Pro	Phe	Arg
5	Phe	His 130	Gln	Phe	Arg	Thr	Leu 135	Lys	Ala	Ala	Val	Gly 140	Val	Ser	Val	Val
	Ile 145	Trp	Ala	Lys	Glu	Leu 150	Leu	Thr	Ser	Ile	Туг 155	Phe	Leu	Met	His	Glu 160
	Glu	Val	Ile	Glu	Asp 165	Glu	Asn	Gln	His	Arg 170	Val	Cys	Phe	Glu	His 175	Tyr
10	Pro	Ile	Gln	Ala 180	Trp	Gln	Arg	Ala	Ile 185	Asn	Tyr	Tyr	Arg	Phe 190	Leu	Val
	Gly	Phe	Leu 195	Phe	Pro	Ile	Cys	Leu 200	Leu	Leu	Ala	Ser	Tyr 205	Gln	Gly	Ile
15	Leu	Arg 210	Ala	Val	Arg	Arg	Ser 215	His	Gly	Thr	Gln	Lys 220	Ser	Arg	Lys	Asp
	Gln 225	Ile	Lys	Arg	Leu	Val 230	Leu	Ser	Thr	Val	Val 235	Ile	Phe	Leu	Ala	Cys 240
	Phe	Leu	Pro	Tyr	His 245	Val	Leu	Leu	Leu	Val 250	Arg	Ser	Val	Trp	Glu 255	
20	Ser	Cys	Asp	Phe 260	Ala	Lys	Gly	Val	Phe 265	Asn	Ala	Tyr	His	Phe 270	Ser	Leu
	Leu	Leu	Thr 275	Ser	Phe	Asn	Cys	Val 280	Ala	Asp	Pro	Val	Leu 285	Tyr	Cys	Phe
25	Val	Ser 290	Glu	Thr	Thr	His	Arg 295	Asp	Leu	Ala	Arg	Leu 300	Arg	Gly	Ala	Cys
	Leu 305	Ala	Phe	Leu	Thr	Cys 310	Ser	Arg	Thr	Gly	Arg 315	Ala	Arg	Glu	Ala	Tyr 320
	Pro	Leu	Gly	Ala	Pro 325	Glu	Ala	Ser	Gly	Lys 330	Ser	Gly	Ala	Gln	Gly 335	Glu
30	Glu	Pro	Glu	Leu 340	Leu	Thr	Lys	Leu	His 345	Pro	Ala	Phe	Gln	Thr 350	Pro	Asn
	Ser	Pro	Gly 355		Gly	Gly	Phe	Pro 360		Gly	Arg	Leu	Ala 365			
	(228) IN	FORM	ATIO	N FO	R SE	Q ID	NO:	227:								
35	(i) SE		CE C					-							

(B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:227:

5	ATGGATATTC	TTTGTGAAGA	AAATACTTCT	TTGAGCTCAA	CTACGAACTC	CCTAATGCAA	6
	TTAAATGATG	ACAACAGGCT	CTACAGTAAT	GACTTTAACT	CCGGAGAAGC	TAACACTTCT	120
	GATGCATTTA	ACTGGACAGT	CGACTCTGAA	AATCGAACCA	ACCTTTCCTG	TGAAGGGTGC	180
	CTCTCACCGT	CGTGTCTCTC	CTTACTTCAT	CTCCAGGAAA	AAAACTGGTC	TGCTTTACTG	240
	ACAGCCGTAG	TGATTATTCT	AACTATTGCT	GGAAACATAC	TCGTCATCAT	GGCAGTGTCC	300
10	CTAGAGAAAA	AGCTGCAGAA	TGCCACCAAC	TATTTCCTGA	TGTCACTTGC	CATAGCTGAT	360
	ATGCTGCTGG	GTTTCCTTGT	CATGCCCGTG	TCCATGTTAA	CCATCCTGTA	TGGGTACCGG	420
	TGGCCTCTGC	CGAGCAAGCT	TTGTGCAGTC	TGGATTTACC	TGGACGTGCT	CTTCTCCACG	480
	GCCTCCATCA	TGCACCTCTG	CGCCATCTCG	CTGGACCGCT	ACGTCGCCAT	CCAGAATCCC	540
	ATCCACCACA	GCCGCTTCAA	CTCCAGAACT	AAGGCATTTC	TGAAAATCAT	TGCTGTTTGG	600
15	ACCATATCAG	TAGGTATATC	CATGCCAATA	CCAGTCTTTG	GGCTACAGGA	CGATTCGAAG	660
	GTCTTTAAGG	AGGGGAGTTG	CTTACTCGCC	GATGATAACT	TTGTCCTGAT	CGGCTCTTTT	720
	GTGTCATTTT	TCATTCCCTT	AACCATCATG	GTGATCACCT	ACTTTCTAAC	TATCAAGTCA	780
	CTCCAGAAAG	AAGCTACTTT	GTGTGTAAGT	GATCTTGGCA	CACGGGCCAA	ATTAGCTTCT	840
	TTCAGCTTCC	TCCCTCAGAG	TTCTTTGTCT	TCAGAAAAGC	TCTTCCAGCG	GTCGATCCAT	900
20	AGGGAGCCAG	GGTCCTACAC	AGGCAGGAGG	ACTATGCAGT	CCATCAGCAA	TGAGCAAAAG	960
	GCAAAGAAGG	TGCTGGGCAT	CGTCTTCTTC	CTGTTTGTGG	TGATGTGGTG	CCCTTTCTTC	1020
	ATCACAAACA	TCATGGCCGT	CATCTGCAAA	GAGTCCTGCA	ATGAGGATGT	CATTGGGGCC	1080
	CTGCTCAATG	TGTTTGTTTG	GATCGGTTAT	CTCTCTTCAG	CAGTCAACCC	ACTAGTCTAC	1140
	ACACTGTTCA	ACAAGACCTA	TAGGTCAGCC	TTTTCACGGT	ATATTCAGTG	TCAGTACAAG	1200
25	GAAAACAAAA	AACCATTGCA	GTTAATTTTA	GTGAACACAA	TACCGGCTTT	GGCCTACAAG	1260
	TCTAGCCAAC	TTCAAATGGG	ACAAAAAAAG	AATTCAAAGC	AAGATGCCAA	GACAACAGAT	1320
	AATGACTGCT	CAATGGTTGC	TCTAGGAAAG	CAGTATTCTG	AAGAGGCTTC	TAAAGACAAT	1380
	AGCGACGGAG	TGAATGAAAA	GGTGAGCTGT	GTGTGA			1416

(229)	INFORMATION	FOR	SEQ	ID	NO:228:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 470 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:

- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:228:
- Met Asp Ile Leu Cys Glu Glu Asn Thr Ser Leu Ser Ser Thr Thr Asn

 10 1 5 10 15

 Ser Leu Met Gln Leu Asn Asp Asp Asn Arg Leu Tyr Ser Asn Asp Phe
 20 25 30
 - Asn Ser Gly Glu Ala Asn Thr Ser Asp Ala Phe Asn Trp Thr Val Asp 35 40 45
- Ser Glu Asn Arg Thr Asn Leu Ser Cys Glu Gly Cys Leu Ser Pro Ser 50 55 60
 - Cys Leu Ser Leu Leu His Leu Gln Glu Lys Asn Trp Ser Ala Leu Leu 65 70 75 80
- Thr Ala Val Val Ile Ile Leu Thr Ile Ala Gly Asn Ile Leu Val Ile 20 85 90 95
 - Met Ala Val Ser Leu Glu Lys Lys Leu Gln Asn Ala Thr Asn Tyr Phe 100 105 110
 - Leu Met Ser Leu Ala Ile Ala Asp Met Leu Leu Gly Phe Leu Val Met 115 120 125
- 25 Pro Val Ser Met Leu Thr Ile Leu Tyr Gly Tyr Arg Trp Pro Leu Pro 130 135 140
 - Ser Lys Leu Cys Ala Val Trp Ile Tyr Leu Asp Val Leu Phe Ser Thr 145 150 155 160
- Ala Ser Ile Met His Leu Cys Ala Ile Ser Leu Asp Arg Tyr Val Ala 30 165 170 175
 - Ile Gln Asn Pro Ile His His Ser Arg Phe Asn Ser Arg Thr Lys Ala 180 185 190
 - Phe Leu Lys Ile Ile Ala Val Trp Thr Ile Ser Val Gly Ile Ser Met 195 : 200 205
- Pro Ile Pro Val Phe Gly Leu Gln Asp Asp Ser Lys Val Phe Lys Glu 210 215 220
 - Gly Ser Cys Leu Leu Ala Asp Asp Asn Phe Val Leu Ile Gly Ser Phe 225 230 235 240

													•				
		Val	Ser	Phe	Phe	Ile 245	Pro	Leu	Thr	Ile	Met 250	Val	Ile	Thr	Tyr	Phe 255	Leu
		Thr	Ile	Lys	Ser 260	Leu	Gln	Lys	Glu	Ala 265	Thr	Leu	Cys	Val	Ser 270	Asp	Leu
5		Gly	Thr	Arg 275	Ala	Lys	Leu	Ala	Ser 280	Phe	Ser	Phe	Leu	Pro 285	Gln	Ser	Ser
		Leu	Ser 290	Ser	Glu	Lys	Leu	Phe 295	Gln	Arg	Ser	Ile	His 300	Arg	Glu	Pro	Gly
10		Ser 305	Tyr	Thr	Gly	Arg	Arg 310	Thr	Met	-Gln	Ser	Ile 315	Ser	Asn	Glu	Gln	Lys 320
	. • •	Ala	Lys	Lys	Val	Leu 325	Gly	Ile	Val	Phe	Phe 330	Leu	Phe	Val		Met 335	Trp
		Cys	Pro	Phe	Phe 340	Ile	Thr	Asn	Ile	Met 345	Ala	Val	Ile	Cys	Lys 350	Glu	Ser
15		Cys	Asn	Glu 355	Asp	Val	Ile	Gly	Ala 360	Leu	Leu	Asn	Val	Phe 365	Val	Trp	Ile
٠		_	Tyr 370		Ser	Ser	Ala	Val 375	Asn	Pro	Leu	Val	Tyr 380	Thr	Leu	Phe	Asn
20 -		Lys 385	Thr	Tyr	Arg	Ser	Ala 390		Ser	Arg	Tyr	Ile 395	Gln	Cys :	Gln	Tyr	Lys 400
		Glu	Asn	Lys	Lys	Pro 405	Leu	Gln	Leu	Ile	Leu 410	Val	Asn	Thr	Ile	Pro 415	Ala
		Leu	Ala	Tyr	Lys 420	Ser	Ser	Gln	Leu	Gln 425	Met	Gly	Gln	Lys	Lys 430	Asn	Ser
25		Lys	Gln	Asp 435	Ala	Lys	Thr	Thr	Asp 440	Asn	Asp	Cys	Ser	Met 445	Val	Ala	Leu
		Gly	_	Gln	Tyr	Ser	Glu	Glu 455	Ala	Ser	Lys	Asp	Asn 460		Asp	Gly	Val
30		Asn 465	Glu	Lys	Val	Ser	Cys 470	Val									
	(230)	INI (FORM	ATIO	v FOI	R SE(Q ID	NO:2	229:								
		(i)						ISTIC									
					NGTH PE: 1			ase p acid	pair	S							
35								sina'	ا ۾ ا								

(ii) MOLECULE TYPE: DNA (genomic)

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:229:

	ATGGTGAACC	TGAGGAATGC	GGTGCATTCA	TTCCTTGTGC	ACCTAATTGG	CCTATTGGTT	60
	TGGCAATGTG	ATATTTCTGT	GAGCCCAGTA	GCAGCTATAG	TAACTGACAT	TTTCAATACC	120
	TCCGATGGTG	GACGCTTCAA	ATTCCCAGAC	GGGGTACAAA	ACTGGCCAGC	ACTTTCAATC	180
5	GTCATCATAA	TAATCATGAC	AATAGGTGGC	AACATCCTTG	TGATCATGGC	AGTAAGCATG	240
	GAAAAGAAAC	TGCACAATGC	CACCAATTAC	TTCTTAATGT	CCCTAGCCAT	TGCTGATATG	300
	CTAGTGGGAC	TACTTGTCAT	GCCCCTGTCT	CTCCTGGCAA	TCCTTTATGA	TTATGTCTGG	360
	CCACTACCTA	GATATTTGTG	CCCCGTCTGG	ATTTCTTTAG	ATGTTTTATT	TTCAACAGCG	420
	TCCATCATGC	ACCTCTGCGC	TATATCGCTG	GATCGGTATG	TAGCAATACG	TAATCCTATT	480
10	GAGCATAGCC	GTTTCAATTC	GCGGACTAAG	GCCATCATGA	AGATTGCTAT	TGTTTGGGCA	540
	ATTTCTATAG	GTGTATCAGT	TCCTATCCCT	GTGATTGGAC	TGAGGGACGA	AGAAAAGGTG	600
	TTCGTGAACA	ACACGACGTG	CGTGCTCAAC	GACCCAAATT	TCGTTCTTAT	TGGGTCCTTC	660
	GTAGCTTTCT	TCATACCGCT	GACGATTATG	GTGATTACGT	ATTGCCTGAC	CATCTACGTT	720
	CTGCGCCGAC	AAGCTTTGAT	GTTACTGCAC	GGCCACACCG	AGGAACCGCC	TGGACTAAGT	780
15	CTGGATTTCC	TGAAGTGCTG	CAAGAGGAAT	ACGGCCGAGG	AAGAGAACTC	TGCAAACCCT	840
	AACCAAGACC	AGAACGCACG	CCGAAGAAAG	AAGAAGGAGA	GACGTCCTAG	GGGCACCATG	900
	CAGGCTATCA	ACAATGAAAG	AAAAGCTAAG	AAAGTCCTTG	GGATTGTTTT	CTTTGTGTTT	960
	CTGATCATGT	GGTGCCCATT	TTTCATTACC	AATATTCTGT	CTGTTCTTTG	TGAGAAGTCC	1020
	TGTAACCAAA	AGCTCATGGA	AAAGCTTCTG	AATGTGTTTG	TTTGGATTGG	CTATGTTTGT	1080
20	TCAGGAATCA	ATCCTCTGGT	GTATACTCTG	TTCAACAAAA	TTTACCGAAG	GGCATTCTCC	1140
	AACTATTTGC	GTTGCAATTA	TAAGGTAGAG	AAAAAGCCTC	CTGTCAGGCA	GATTCCAAGA	1200
	GTTGCCGCCA	CTGCTTTGTC	TGGGAGGGAG	CTTAATGTTA	ACATTTATCG	GCATACCAAT	1260
	GAACCGGTGA	TCGAGAAAGC	CAGTGACAAT	GAGCCCGGTA	TAGAGATGCA	AGTTGAGAAT	1320
	TTAGAGTTAC	CAGTAAATCC	CTCCAGTGTG	GTTAGCGAAA	GGATTAGCAG	TGTGTGA	1377

25 (231) INFORMATION FOR SEQ ID NO:230:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 458 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:

- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:230:

	(11	, 52	, LUDE							_						
5	Met 1	Val	Asn	Leu	Arg 5	Asn	Ala	Val	His	Ser 10	Phe	Leu	Val	His	Leu 15	Ile
	Gly	Leu	Leu	Val 20	Trp	Gln	Cys	Asp	Ile 25	Ser	Val	Ser	Pro	Val 30	Ala	Ala
	Ile	Val	Thr 35	Asp	Ile	Phe	Asn	Thr 40	Ser	Asp	Gly	Gly	Arg 45	Phe	Lys	Phe
10	Pro	Asp 50	Gly	Val	Gln	Asn	Trp 55	Pro	Ala	Leu	Ser	Ile 60	Val	Ile	Ile	Ile
	Ile 65	Met	Thr	Ile	Gly	Gly 70	Asn	Ile	Leu	Val	Ile 75	Met	Ala	Val	Ser	Met 80
15	Glu	Lys	Lys	Leu	His 85	Asn	Ala	Thr	Asn	Tyr 90	Phe	Leu	Met	Ser	Leu 95	Ala
	Ile	Ala	Asp	Met 100	Leu	Val	Gly	Leu	Leu 105	Val	Met	Pro	Leu	Ser 110	Leu	Leu
	Ala	Ile	Leu 115	Tyr	Asp	Tyr	Val	Trp 120	Pro	Leu	Pro	Arg	Tyr 125	Leu	Cys	Pro
20	Val	Trp 130	Ile	Ser	Leu	Asp	Val 135	Leu	Phe	Ser	Thr	Ala 140	Ser	Ile	Met	His
	Leu 145	Cys	Ala	Ile	Ser	Leu 150	Asp	Arg	Tyr	Val	Ala 155	Ile	Arg	Asn	Pro	Ile 160
25	Glu	His	Ser	Arg	Phe 165	Asn	Ser	Arg	Thr	Lys 170	Ala	Ile	Met	Lys	Ile 175	Ala
	Ile	Val	Trp	Ala 180	Ile	Ser	Ile	Gly	Val 185		Val	Pro	Ile	Pro 190	Val	Ile
	Gly	Leu	Arg 195	Asp	Glu	Glu	Lys	Val 200		Val	Asn	Asn	Thr 205		Cys	Val
30	Leu	Asn 210		Pro	Asn	Phe	Val 215		Ile	Gly	Ser	Phe 220		Ala	Phe	Phe
	11e 225		Leu	Thr	Ile	Met 230		Ile	Thr	Tyr	Cys 235		Thr	Ile	Tyr	Val 240
35	Leu	Arg	Arg	Gln	Ala 245		Met	Leu		His 250		His	Thr	Glu	Glu 255	Pro
	Pro	Gly	Leu	Ser	Lev	Asp	Phe	Leu	Lys	Cys	Cys	Lys	Arg	J Asn	Thr	Ala

					260					265					270			
		Glu	Glu	Glu 275	Asn	Ser	Ala	Asn	Pro 280	Asn	Gln	Asp	Gln	Asn 285	Ala	Arg	Arg	
5		Arg	Lys 290	Lys	Lys	Glu	Arg	Arg 295	Pro	Arg	Gly	Thr	Met 300	Gln	Ala	Ile	Asn	
	4	Asn 305	Glu	Arg	Lys	Ala	Lys 310	Lys	Val	Leu	Gly	Ile 315	Val	Phe	Phe	Val	Phe 320	
		Leu	Ile	Met	Trp	Cys 325	Pro	Phe	Phe	Ile	Thr 330	Asn	Ile	Leu	Ser	Val 335	Leu	
10		Cys	Glu	Lys	Ser 340	Cys	Asn	Gln	Lys	Leu 345	Met	Glu	Lys	Leu	Leu 350	Asn	Val	
		Phe	Val	Trp 355	Ile	Gly	Tyr	Val	Cys 360	Ser	Gly	Ile	Asn	Pro 365	Leu	Val	Tyr	
15		Thr	Leu 370	Phe	Asn	Lys	Ile	Tyr 375	Arg	Arg	Ala	Phe	Ser 380	Asn	Tyr	Leu	Arg	
	a.	Cys 385	Asn	Tyr	Lys	Val	Glu 390	Lys	Lys	Pro	Pro	Val 395	Arg	Gln	Ile	Pro	Arg 400	
	ċ	Val	Ala	Ala	Thr	Ala 405	Leu	Ser	Gly	Arg	Glu 410	Leu	Asn	Val	Asn	Ile 415	Tyr	
20		Arg	His	Thr	Asn 420	Glu	Pro	Val	Ile	Glu 425	Lys	Ala	Ser	Asp	Asn 430	Glu	Pro	
		Gly	Ile	Glu 435	Met	Gln	Val	Glu	Asn 440	Leu	Glu	Leu	Pro	Val 445	Asn	Pro	Ser	
25		Ser	Val 450	Val	Ser	Glu	Arg	Ile 455	Ser	Ser	Val							
	(232) IN	FORM	OITA	N FOI	R SE	Q ID	NO:2	231:									
		(i)	(A)	LEI TY	CE CI NGTH PE: 1	: 10 nucl	68 ba	ase pacid	pair	5			•					
30					RANDI POLO			_	le				•					
		(i:	i) M	OLEC	ULE !	TYPE	: DN	A (g	enom	ic)								
		(x:	i) S	EQUE	NCE 1	DESC	RIPT	ION:	SEQ	ID 1	NO:2	31:						
	ATGG.	ATCA	GT T	CCCT	GAAT(C AG	TGAC	AGAA	AAC'	rttg:	AGT A	ACGA'	rgat:	rt G	GCTG:	AGGC	2	60
35	TGTT.	ATAT'	TG G	GGAC	ATCG'	r gg	TCTT'	rggg	ACT	GTGT	TCC '	rgtc	CATA:	rt c	ract(CCGT	2 3	L20
	ATCT	TTGC	CA T	TGGC	CTGG'	r gg	GAAA'	TTTG	TTG	GTAG'	TGT '	TTGC	CCTC	AC C	AACA	GCAA(3 3	180

WO 00/22129 PCT/US99/23938

199

	AAGCCCAAGA GTGTCACCGA	CATTTACCTC	CTGAACCTGG	CCTTGTCTGA	TCTGCTGTTT	240
	GTAGCCACTT TGCCCTTCTG	GACTCACTAT	TTGATAAATG	AAAAGGGCCT	CCACAATGC	300
	ATGTGCAAAT TCACTACCGC	CTTCTTCTTC	ATCGGCTTTT	TTGGAAGCAT	ATTCTTCATO	360
	ACCGTCATCA GCATTGATAG	GTACCTGGCC	ATCGTCCTGG	CCGCCAACTC	CATGAACAAC	420
5	CGGACCGTGC AGCATGGCGT	CACCATCAGC	CTAGGCGTCT	GGGCAGCAGC	CATTTTGGT	480
	GCAGCACCCC AGTTCATGTT	CACAAAGCAG	AAAGAAAATG	AATGCCTTGG	TGACTACCCC	540
	GAGGTCCTCC AGGAAATCTG	GCCCGTGCTC	CGCAATGTGG	AAACAAATTT	TCTTGGCTTC	600
	CTACTCCCCC TGCTCATTAT	GAGTTATTGC	TACTTCAGAA	TCATCCAGAC	GCTGTTTTCC	660
	TGCAAGAACC ACAAGAAAGC	CAAAGCCAAG	AAACTGATCC	TTCTGGTGGT	CATCGTGTTT	720
10	TTCCTCTTCT GGACACCCTA	CAACGTTATG	ATTTTCCTGG	AGACGCTTAA	GCTCTATGAC	780
٠.	TTCTTTCCCA GTTGTGACAT	GAGGAAGGAT	CTGAGGCTGG	CCCTCAGTGT	GACTGAGACG	840
	GTTGCATTTA GCCATTGTTG	CCTGAATCCT	CTCATCTATG	CATTTGCTGG	GGAGAAGTTC	900
	AGAAGATACC TTTACCACCT	GTATGGGAAA	TGCCTGGCTG	TCCTGTGTGG	GCGCTCAGTC	960
. :	CACGTTGATT TCTCCTCATC	rgaatcacaa	AGGAGCAGGC	ATGGAAGTGT	TCTGAGCAGC	1020
15	AATTTTACTT ACCACACGAG	rgatggagat	GCATTGCTCC	TTCTCTGA	. ,	1068
	(233) INFORMATION FOR S	SEQ ID NO:2	32:			
	(i) SEQUENCE CHAI (A) LENGTH: 3					ast.
20	(B) TYPE: am: (C) STRANDEDI	ino acid				
	(D) TOPOLOGY		ant	•	•	
	(ii) MOLECULE TYP	PE: protein				
	(xi) SEQUENCE DES	SCRIPTION:	SEQ ID NO:2	32:		
25	Met Asp Gln Phe Pi 1 5	ro Glu Ser	Val Thr Glu 10	Asn Phe Gl	u Tyr Asp . 15	Asp
	Leu Ala Glu Ala Cy 20	ys Tyr Ile	Gly Asp Ile 25	Val Val Ph	e Gly Thr	Val
	Phe Leu Ser Tle D	ne Tyr Ser	Val Tle Dhe	Ala Tle Cl	v Len Val	G) v

	6	55					70.		,	• '	,	75			٠.		80
	V	/al	Ala	Thr	Leu	Pro 85	Phe	Trp	Thr	His	Tyr 90	Leu	Ile	: Asn	ı Glu	Lys 95	Gly
5	I	Leu	His	Asn	Ala 100	Met	Cys	Lys	Phe	Thr 105		Ala	Phe	Phe	Phe 110		Gly
	F	he	Phe	Gly 115	Ser	Ile	Phe	Phe	Ile 120		Val	Ile	Ser	Ile 125		Arg	Tyr
	I	eu	Ala 130		Val	Leu	Ala	Ala 135		Ser	Met	Asn	Asn 140		Thr	Val	Gln
10		(i.s .45	Glÿ	Val	Thr	Ile	Ser 150	Leu	Gly	Val	Trp	Ala 155	Ala	Ala	Ile	Leu	Val 160
	A	la	Ala	Pro	Gln	Phe 165	Met	Phe	Thr	Lys	Gln 170	Lys	Glu	Asn	Glu	Cys 175	Leu
15	G	ly	Asp	Tyr	Pro 180	Glu	Val	Leu	Gln	Glu 185	Ile	Trp	Pro	Val	Leu 190	Arg	Asn
	v	al	Glu	Thr 195	Asn	Phe	Leu	Gly	Phe 200	Leu	Leu	Pro	Leu	Leu 205	Ile	Met	Ser
	T	yr	Cys 210	Tyr	Phe	Arg	Ile	Ile 215	Gln	Thr	Leu	Phe	Ser 220	Cys	Lys	Asn	His
20	L 2	ys 25	Lys	Ala	Lys	Ala	Lys 230	Lys	Leu	Ile	Leu	Leu 235	Val	Val	Ile	Val	Phe 240
	P	he	Leu	Phe	Trp	Thr 245	Pro	Tyr	Asn	Val	Met 250	Ile	Phe	Leu	Glu	Thr 255	Leu
25	L	ys	Leu	Tyr	Asp 260	Phe	Phe	Pro	Ser	Cys 265	Asp	Met	Arg	Lys	Asp 270	Leu	Arg
	L	eu	Ala	Leu 275	Ser	Val	Thr	Glu	Thr 280	Val	Ala	Phe	Ser	His 285	Cys	Cys	Leu
	A:	sn	Pro 290	Leu	Ile	Tyr	Ala	Phe 295	Ala	Gly	Glu	Lys	Phe 300	Arg	Arg	Tyr	Leu
30	T;	yr 05	His	Leu	Tyr	Gly	Lys 310	Cys	Leu	Ala	Val	Leu 315	Cys	Gly	Arg	Ser	Val 320
	H	is	Val	Asp		Ser 325	Ser	Ser	Glu	Ser	Gln 330	Arg	Ser	Arg	His	Gly 335	Ser
35	Vā	al	Leu	Ser	Ser 340	Asn	Phe	Thr	Tyr	His 345	Thr	Ser	Asp	Gly	Asp 350	Ala	Leu
	Le	eu	Leu	Leu 355													

RNSDOCID <WO 00212941 145

	(234)	INFORMATION TON BEY IT HOLDED!	
5		 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 29 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
		(ii) MOLECULE TYPE: DNA (genomic)	
		(iv) ANTI-SENSE: NO	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO:233:	
10	GGCTT	AAGAG CATCATCGTG GTGCTGGTG	29
	(235)	INFORMATION FOR SEQ ID NO:234:	
15		(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 34 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA (genomic)	
		(iv) ANTI-SENSE: YES	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO:234:	
20	GTCAC	CACCA GCACCACGAT GATGCTCTTA AGCC	34
	(236)	INFORMATION FOR SEQ ID NO:235:	
25		 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	. •
		(ii) MOLECULE TYPE: DNA (genomic) (xi) SEQUENCE DESCRIPTION: SEQ ID NO:235:	
	CAAAG	BAAAGT ACTGGGCATC GTCTTCTTCC T	31
30	(237)	INFORMATION FOR SEQ ID NO:236:	
35		 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
<i></i>		(ii) MOLECULE TYPE: DNA (genomic)	

	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:236:	
	TGCTCTAGAT TCCAGATAGG TGAAAACTTG	30
	(238) INFORMATION FOR SEQ ID NO.237:	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 50 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
10	(iv) ANTI-SENSE: NO	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:237:	
	CTAGGGGCAC CATGCAGGCT ATCAACAATG AAAGAAAAGC TAAGAAAGTC	50
	(239) INFORMATION FOR SEQ ID NO:238:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 50 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
20	(iv) ANTI-SENSE: YES	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:238:	
	CAAGGACTTT CTTAGCTTTT CTTTCATTGT TGATAGCCTG CATGGTGCCC	50
	(240) INFORMATION FOR SEQ ID NO:239:	
25	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 35 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
30	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:239:	
	CGGCGGCAGA AGGCGAAACG CATGATCCTC GCGGT	35
	(241) INFORMATION FOR SEQ ID NO:240:	
35	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 35 base pairs (B) TYPE: nucleic acid	

BNSDOCID -WO 00212941 14-

	(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:240:	
5	ACCGCGAGGA TCATGCGTTT CGCCTTCTGC CGCCG	. 35
	(242) INFORMATION FOR SEQ ID NO:241:	
10	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:241:	
	GAGACATATT ATCTGCCACG GAGG	24
15	(243) INFORMATION FOR SEQ ID NO:242:	
20	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:242:	
	TTGGCATAGA AACCGGACCC AAGG	24
	(244) INFORMATION FOR SEQ ID NO:243:	
25	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 28 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
30	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:243:	
	PAAGAATTCC ATAAAAATTA TGGAATGG	. 28
	(245) INFORMATION FOR SEQ ID NO:244:	
	(i) SEOUENCE CHARACTERISTICS:	

	(A) LENGTH: 30 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
5	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:244:	
	CCAGGATCCA GCTGAAGTCT TCCATCATTC	30
	(246) INFORMATION FOR SEQ ID NO:245:	
10	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1071 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
15	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:245:	
	ATGAATGGGG TCTCGGAGGG GACCAGAGGC TGCAGTGACA GGCAACCTGG GGTCCTGACA	60
٠.	CGTGATCGCT CTTGTTCCAG GAAGATGAAC TCTTCCGGAT GCCTGTCTGA GGAGGTGGGG	120
	TCCCTCCGCC CACTGACTGT GGTTATCCTG TCTGCGTCCA TTGTCGTCGG AGTGCTGGGC	180
	AATGGGCTGG TGCTGTGGAT GACTGTCTTC CGTATGGCAC GCACGGTCTC CACCGTCTGC	240
20	TTCTTCCACC TGGCCCTTGC CGATTTCATG CTCTCACTGT CTCTGCCCAT TGCCATGTAC	300
	TATATTGTCT CCAGGCAGTG GCTCCTCGGA GAGTGGGCCT GCAAACTCTA CATCACCTTT	360
	GTGTTCCTCA GCTACTTTGC CAGTAACTGC CTCCTTGTCT TCATCTCTGT GGACCGTTGC	420
	ATCTCTGTCC TCTACCCCGT CTGGGCCCTG AACCACCGCA CTGTGCAGCG GGCGAGCTGG	480
	CTGGCCTTTG GGGTGTGGCT CCTGGCCGCC GCCTTGTGCT CTGCGCACCT GAAATTCCGG	540
25	ACAACCAGAA AATGGAATGG CTGTACGCAC TGCTACTTGG CGTTCAACTC TGACAATGAG	600
	ACTGCCCAGA TTTGGATTGA AGGGGTCGTG GAGGGACACA TTATAGGGAC CATTGGCCAC	660
	TTCCTGCTGG GCTTCCTGGG GCCCTTAGCA ATCATAGGCA CCTGCGCCCA CCTCATCCGG	720
	GCCAAGCTCT TGCGGGAGGG CTGGGTCCAT GCCAACCGGC CCGCGAGGCT GCTGCTGGTG	780
	CTGGTGAGCG CTTTCTTAT CTTCTGGTCC CCGTTTAACG TGGTGCTGTT GGTCCATCTG	840
30	TGGCGACGGG TGATGCTCAA GGAAATCTAC CACCCCGGA TGCTGCTCAT CCTCCAGGCT	900
	AGCTTTGCCT TGGGCTGTGT CAACAGCAGC CTCAACCCCT TCCTCTACGT CTTCGTTGGC	960

AGAGATTTCC AAGAAAAGTT TTTCCAGTCT TTGACTTCTG CCCTGGCGAG GGCGTTTGGA 1020

	GAGGA	.GGAG	т тт	CTGI	CATC	CTG	TCCC	CGT:	GGCA	ACGC	cc c	CCGG	GAAT	G A			1071
	(247) INFORMATION FOR SEQ ID NO:246:																
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 356 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: (D) TOPOLOGY: not relevant 																
	(ii) MOLECULE TYPE: protein																
10	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:246:																
		Met 1	Asn	Gly	Val	Ser 5	Glu	Gly	Thr	Arg	Gly 10	Cys	Ser	Asp	Arg	Gln 15	Pro
		Gly	Val	Leu	Thr 20	Arg	Asp	Arg	Ser	Cys 25	Ser	Arg	Lys	Met	Asn 30	Ser	Ser
15		Gly	Cys	Leu 35	Ser	Glu	Glu	Val	Gly 40	Ser	Leu	Arg	Pro	Leu 45	Thr	Val	Val
		Ile	Leu 50	Ser	Ala	Ser	Ile	Val 55	Val	Gly	Val	Leu	Gly 60	Asn	Gly	Leu	Val
20		Leu 65	Trp	Met	Thr	Val	Phe 70	Arg	Met	Ala	Arg	Thr 75	Val	Ser	Thr	Val	Cys 80
		Phe	Phe	His	Leu	Ala 85	Leu	Ala	Asp	Phe	Met 90	Leu	Ser	Leu	Ser	Leu 95	Pro
		Ile	Ala	Met	Tyr 100	Tyr	Ile	Val	Ser	Arg 105	Gln	Trp	Leu	Leu	Gly 110	Glu	Trp
25		Ala	Cys	Lys 115	Leu	Tyr	Ile	Thr	Phe 120	Val	Phe	Leu	Ser	Tyr 125	Phe	Ala	Ser
		Asn	Cys 130	Leu	Leu	Val	Phe	Ile 135		Val	Asp	Arg	Cys 140	Ile	Ser	Val	Leu
30		Tyr 145	Pro	Val	Trp	Ala	Leu 150	Asn	His	Arg	Thr	Val 155	Gln	Arg	Ala	Ser	Trp 160
		Leu	Ala	Phe	Gly	Val 165	Trp	Leu	Leu	Ala	Ala 170	Ala	Leu	Cys	Ser	Ala 175	His
		Leu	Lys	Phe	Arg 180	Thr	Thr	Arg	Lys	Trp 185	Asn	Gly	Cys	Thr	His 190	Cys	Tyr
35		Leu	Ala	Phe 195		Ser	Asp	Asn	Glu 200		Ala	Gln	Ile	Trp 205		Glu	Gly

	Va	11 Val 210		Gly	His	Ile	Ile 215	Gly	Thr	Ile	Gly	His 220	Phe	Leu	Leu	Gly	
	Ph 22	e Leu 5	Gly	Pro	Leu	Ala 230	Ile	Ile	Gly	Thr	Cys 235	Ala	His	Leu	Ile	Arg 240	
5	Al	a Lys	Leu	Leu	Arg 245	Glu	Gly	Trp	Val	His 250	Ala	Asn	Arg	Pro	Ala 255	Arg	
	Le	u Leu	Leu	Val 260	Leu	Val	Ser	Ala	Phe 265	Phe	Ile	Phe	Trp	Ser 270	Pro	Phe	
10	As	n Val	Val 275	Leu	Leu	Val	His	Leu 280	Trp	Arg	Arg	Val	Met 285	Leu	Lys	Glu	
	11	e Tyr 290	His	Pro	Arg	Met	Leu 295	Leu	Ile	Leu	Gln	Ala 300	Ser	Phe	Ala	Leu	
	30					310					315					320	
15		g Asp			325					330					335		
		g Ala	-	340	Glu	Glu	Glu	Phe	Leu 345	Ser	Ser	Cys	Pro	Arg 350	Gly	Asn	
20		a Pro	355						•								
	(248) II																
25	. (:	(B) (C)	LEN TYP STR	CE CH IGTH: PE: n RANDE POLOG	32 ucle DNES	base ic a S: s	pai cid ingl	rs									
	(:	Li) MC	LECU	LE T	YPE:	DNA	. (ge	nomi	c)								
	. (2	ci) SE	QUEN	CE D	ESCR	IPTI	ON:	SEQ	ID N	0:24	7:						
30	GCAGAATT															3:	2
35		(A) (B) (C)	UENC LEN TYP STR		ARAC 30 ucle DNES	TERI base ic a S: s	STIC pai cid ingl	S: rs								·	
	i)	i) MC	LECU	LE T	YPE:	DNA	(ge	nomi	c)								
	(x	i) SE	QUEN	CE D	ESCR	IPTI	ON:	SEQ	ID N	0:24	8:						

WO 00/22129 PCT/US99/23938

	GCTGGATCCC CCGAGCAGTG GCGTTACTTC	30												
	(250) INFORMATION FOR SEQ ID NO:249:	•												
5	(C) STRANDEDNESS: single (D) TOPOLOGY: linear													
	(ii) MOLECULE TYPE: DNA (genomic)													
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:249:													
0	ATGGACCTGC CCCCGCAGCT CTCCTTCGGC CTCTATGTGG CCGCCTTTGC GCTGGGCTTC	60												
	CCGCTCAACG TCCTGGCCAT CCGAGGCGCG ACGGCCCACG CCCGGCTCCG TCTCACCCCT	120												
	AGCCTGGTCT ACGCCCTGAA CCTGGGCTGC TCCGACCTGC TGCTGACAGT CTCTCTGCCC	180												
	CTGAAGGCGG TGGAGGCGCT AGCCTCCGGG GCCTGGCCTC TGCCGGCCTC GCTGTGCCCC	240												
	GTCTTCGCGG TGGCCCACTT CTTCCCACTC TATGCCGGCG GGGGCTTCCT GGCCGCCCTG	300												
15	AGTGCAGGCC GCTACCTGGG AGCAGCCTTC CCCTTGGGCT ACCAAGCCTT CCGGAGGCCG	360												
	TGCTATTCCT GGGGGGTGTG CGCGGCCATC TGGGCCCTCG TCCTGTGTCA CCTGGGTCTG	420												
	GTCTTTGGGT TGGAGGCTCC AGGAGGCTGG CTGGACCACA GCAACACCTC CCTGGGCATC	480												
	AACACACCGG TCAACGGCTC TCCGGTCTGC CTGGAGGCCT GGGACCCGGC CTCTGCCGGC	540												
	CCGGCCCGCT TCAGCCTCTC TCTCCTGCTC TTTTTTCTGC CCTTGGCCAT CACAGCCTTC	600												
20	TGCTACGTGG GCTGCCTCCG GGCACTGGCC CGCTCCGGCC TGACGCACAG GCGGAAGCTG	660												
	CGGGCCGCCT GGGTGGCCGG CGGGGCCCTC CTCACGCTGC TGCTCTGCGT AGGACCCTAC	720												
	AACGCCTCCA ACGTGGCCAG CTTCCTGTAC CCCAATCTAG GAGGCTCCTG GCGGAAGCTG	780												
	GGGCTCATCA CGGGTGCCTG GAGTGTGGTG CTTAATCCGC TGGTGACCGG TTACTTGGGA	840												
	AGGGGTCCTG GCCTGAAGAC AGTGTGTGCG GCAAGAACGC AAGGGGGCAA GTCCCAGAAG	900												
25	TAA	903												
	(251) INFORMATION FOR SEQ ID NO:250:													
30	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 300 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: (D) TOPOLOGY: not relevant 													

	<pre>(ii) MOLECULE TYPE: protein (xi) SEQUENCE DESCRIPTION: SEQ ID NO:250: Met Asp Leu Pro Pro Gln Leu Ser Phe Gly Leu Tyr Val Ala Ala Phe</pre>															
	Met 1	Asp	Leu	Pro	Pro 5	Gln	Leu	Ser	Phe	Gly 10	Leu	Tyr	Val	Ala	Ala 15	Phe
5	Ala	Leu	Gly	Phe 20	Pro	Leu	Asn	Val	Leu 25	Ala	Ile	Arg	Gly	Ala 30	Thr	Ala
	His	Ala	Arg 35	Leu	Arg	Leu	Thr	Pro 40	Ser	Leu	Val	Tyr	Ala 45	Leu	Asn	Leu
10	Gly	Cys 50	Ser	Asp	Leu	Leu	Leu 55	Thr	Val	Ser	Leu	Pro 60	Leu	Lys	Ala	Val
	65		Leu			70					75					80
			Ala		85					90					95	
15				100					105					110		Leu
		-	115					120					125			Ala
20	is .	130					135					140				Leu
	145		Pro			150					155					160
			Pro		165					170					175	
25				180					185					190		Phe
			195					200					205			Ala
30		210					215					220				Trp
	225					230					235					Tyr 240
25					245					250					255	
35	Trp	Arg	г'ns	Leu 260	_	ьen	rte	inr	265		rrb	ser	val	270		Asn

Pro Leu Val Thr Gly Tyr Leu Gly Arg Gly Pro Gly Leu Lys Thr Val

WO 00/22129 PCT/US99/23938

	275 280 285	
	Cys Ala Ala Arg Thr Gln Gly Gly Lys Ser Gln Lys 290 295 300	
	(252) INFORMATION FOR SEQ ID NO:251:	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
10	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:251:	
	CTCAAGCTTA CTCTCTCA CCAGTGGCCA C	31
	(253) INFORMATION FOR SEQ ID NO:252:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	ar.
	(ii) MOLECULE TYPE: DNA (genomic)	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:252:	
	CCCTCCTCCC CCGGAGGACC TAGC	24
	(254) INFORMATION FOR SEQ ID NO:253:	4
25	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1041 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:253:	
30	ATGGATACAG GCCCCGACCA GTCCTACTTC TCCGGCAATC ACTGGTTCGT CTTCTCGGTG	60
	TACCTTCTCA CTTTCCTGGT GGGGCTCCCC CTCAACCTGC TGGCCCTGGT GGTCTTCGTG	120
	GGCAAGCTGC AGCGCCGCCC GGTGGCCGTG GACGTGCTCC TGCTCAACCT GACCGCCTCG	180
	GACCTGCTCC TGCTGCTGTT CCTGCCTTTC CGCATGGTGG AGGCAGCCAA TGGCATGCAC	240
	TGGCCCCTGC CCTTCATCCT CTGCCCACTC TCTGGATTCA TCTTCTTCAC CACCATCTAT	300

	360
	CTGTGGTACA AGACCCGGCC GAGGCTGGGG CAGGCAGGTC TGGTGAGTGT GGCCTGCTGG 420
	CTGTTGGCCT CTGCTCACTG CAGCGTGGTC TACGTCATAG AATTCTCAGG GGACATCTCC 480
	CACAGCCAGG GCACCAATGG GACCTGCTAC CTGGAGTTCC GGAAGGACCA GCTAGCCATC 540
5	CTCCTGCCCG TGCGGCTGGA GATGGCTGTG GTCCTCTTTG TGGTCCCGCT GATCATCACC 600
	AGCTACTGCT ACAGCCGCCT GGTGTGGATC CTCGGCAGAG GGGGCAGCCA CCGCCGGCAG 660
	AGGAGGGTGG CGGGGCTGTT GGCGGCCACG CTGCTCAACT TCCTTGTCTG CTTTGGGCCC 720
	TACAACGTGT CCCATGTCGT GGGCTATATC TGCGGTGAAA GCCCGGCATG GAGGATCTAC 780
	GTGACGCTTC TCAGCACCCT GAACTCCTGT GTCGACCCCT TTGTCTACTA CTTCTCCTCC 840
10	TCCGGGTTCC AAGCCGACTT TCATGAGCTG CTGAGGAGGT TGTGTGGGCT CTGGGGCCAG 900
	TGGCAGCAGG AGAGCAGCAT GGAGCTGAAG GAGCAGAAGG GAGGGGAGGA GCAGAGAGCG 960
	GACCGACCAG CTGAAAGAAA GACCAGTGAA CACTCACAGG GCTGTGGAAC TGGTGGCCAG 1020
	GTGGCCTGTG CTGAAAGCTA G
•	(255) INFORMATION FOR SEQ ID NO:254:
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 346 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: (D) TOPOLOGY: not relevant
20	(ii) MOLECULE TYPE: protein
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:254:
	Met Asp Thr Gly Pro Asp Gln Ser Tyr Phe Ser Gly Asn His Trp Phe 1 5 10 15
25	Val Phe Ser Val Tyr Leu Leu Thr Phe Leu Val Gly Leu Pro Leu Asn 20 25 30
	Leu Leu Ala Leu Val Val Phe Val Gly Lys Leu Gln Arg Arg Pro Val 35 40 45
	Ala Val Asp Val Leu Leu Leu Asn Leu Thr Ala Ser Asp Leu Leu Leu 50 55 60
30	Leu Leu Phe Leu Pro Phe Arg Met Val Glu Ala Ala Asn Gly Met His 65 70 75 80
	Trp Pro Leu Pro Phe Ile Leu Cys Pro Leu Ser Gly Phe Ile Phe Pho

	נ	Chr	Thr	Ile	Tyr 100	Leu	Thr	Ala	Leu	Phe 105	Leu	Ala	Ala	Val	Ser 110	Ile	Gli
	P	arg	Phe	Leu 115	Ser	Val	Ala	His	Pro 120	Leu	Trp	Tyr	Lys	Thr 125	Arg	Pro	Arg
5	I	eu	Gly 130	Gln	Ala	Gly	Leu	Val 135	Ser	Val	Ala	Cys	Trp 140	Leu	Leu	Ala	Ser
		1a .45	His	Cys	Ser	Val	Val 150	Tyr	Val	Ile	Glu	Phe 155	Ser	Gly	Asp	Ile	Ser 160
10	H	lis	Ser	Gln	Gly	Thr 165	Asn	Gly	Thr	Cys	Tyr 170	Leu	Glu	Phe	Arg	Lys 175	Asp
	G	ln	Leu	Ala	Ile 180	Leu	Leu	Pro	Val	Arg 185	Leu	Glu	Met	Ala	Val 190	Val	Leu
	P	he	Val	Val 195	Pro	Leu	Ile	Ile	Thr 200	Ser	Tyr	Cys	Tyr	Ser 205	Àrg	Leu	Val
15	Т		Ile 210	Leu	Gly	Arg	Gly	Gly 215	Ser	His	Arg	Arg	Gln 220	Arg	Arg	Val	Ala
		1y 25	Leu	Leu	Ala	Ala	Thr 230	Leu	Leu	Asn	Phe	Leu 235	Val	Cys	Phe	Gly	Pro 240
20	Т	yr	Asn	Val	Ser	His 245	Val	Val	Gly	Tyr	Ile 250	Cys	Gly	Glu	Ser	Pro 255	Ala
	T	rp	Arg	Ile	Tyr 260	Val	Thr	Leu	Leu	Ser 265	Thr	Leu	Asn	Ser	Cys 270	Val	Asp
	P	ro	Phe	Val 275	Tyr	Tyr	Phe	Ser	Ser 280	Ser	Gly	Phe	Gln	Ala 285	Asp	Phe	His
25	·G		Leu 290	Leu	Arg	Arg	Leu	Cys 295	Gly	Leu	Trp	Gly	Gln 300	Trp	Gln	Gln	Glu
		er 05	Ser	Met	Glu	Leu	Lys 310	Glu	Gln	Lys	Gly	Gly 315	Glu	Glu	Gln	Arg	Ala 320
30	A	sp .	Arg	Pro	Ala	Glu 325	Arg	Lys	Thr	Ser	Glu 330	His	Ser	Gln	Gly	Cys 335	Gly
	т	hr	Gly	Gly	Gln 340	Val	Ala	Cys	Ala	Glu 345	Ser						
	(256)	INF	ORMA	TION	FOR	SEO	ID	NO:2	55:								

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs

 - (B) TYPE: nucleic acid(C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

	(11) MOLECULE TIPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:255:	
	TTTAAGCTTC CCCTCCAGGA TGCTGCCGGA C	31
	(257) INFORMATION FOR SEQ ID NO:256:	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: not relevant 	
10	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:256:	
	GGCGAATTCT GAAGGTCCAG GGAAACTGCT A	31
	(258) INFORMATION FOR SEQ ID NO:257:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 993 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:257:	
	ATGCTGCCGG ACTGGAAGAG CTCCTTGATC CTCATGGCTT ACATCATCAT CTTCCTCACT	60
	GGCCTCCCTG CCAACCTCCT GGCCCTGCGG GCCTTTGTGG GGCGGATCCG CCAGCCCCAG	120
	CCTGCACCTG TGCACATCCT CCTGCTGAGC CTGACGCTGG CCGACCTCCT CCTGCTGCTG	180
	CTGCTGCCCT TCAAGATCAT CGAGGCTGCG TCGAACTTCC GCTGGTACCT GCCCAAGGTC	240
25	GTCTGCGCCC TCACGAGTTT TGGCTTCTAC AGCAGCATCT ACTGCAGCAC GTGGCTCCTG	300
	GCGGGCATCA GCATCGAGCG CTACCTGGGA GTGGCTTTCC CCGTGCAGTA CAAGCTCTCC	360
	CGCCGGCCTC TGTATGGAGT GATTGCAGCT CTGGTGGCCT GGGTTATGTC CTTTGGTCAC	420
	TGCACCATCG TGATCATCGT TCAATACTTG AACACGACTG AGCAGGTCAG AAGTGGCAAT	480
	GAAATTACCT GCTACGAGAA CTTCACCGAT AACCAGTTGG ACGTGGTGCT GCCCGTGCGG	540
30	CTGGAGCTGT GCCTGGTGCT CTTCTTCATC CCCATGGCAG TCACCATCTT CTGCTACTGG	600
	CGTTTTGTGT GGATCATGCT CTCCCAGCCC CTTGTGGGGG CCCAGAGGCG GCGCCGAGCC	660
	GTGGGGCTGG CTGTGGTGAC GCTGCTCAAT TTCCTGGTGT GCTTCGGACC TTACAACGTG	720

'WO 00/22129

	TCCCACC	rgg T	GGGG'	ratc:	A CCI	AGAG	AAAA	AGC	CCT	GT (GCG	STCA	A TA	GCCG'	rggt	•	780
	TTCAGTT	CAC T	CAAC	GCCA	G TC	rgga	ccc	CTG	TCT	CT A	ATTT(CTCT:	rc T	rcag:	rggt	3 8	340
	CGCAGGG	CAT T	TGGG	AGAGO	G GCT	rgcao	GTG	CTG	CGGA	ATC A	AGGG	TCC.	rc co	CTGT:	rgggz	A 9	900
	CGCAGAG	CA A	AGAC	ACAG	C AG	AGGG	GACA	AATO	AGG	ACA (GGG:	rgrgo	G T	CAAG	GAGA	A 9	960
5	GGGATGC	CAA G	TTCG	GACT:	CA(CTAC	AGAG	TAG								9	993
	(259) II	1FORM	ATIO	N FOI	R SE	O ID	NO:	258:									
10		(B (C) LEI) TY!) ST!) TO!	NGTH PE: 8 RANDI POLO	: 362 amino EDNES GY: 1	2 am: o ac: SS: not :	ino a id relev	acids vant	3								
		·				-			TD .	10 · 01	- 0 -			•	•		
		ki) S	-	,									3 3-		-1 -		
15	Met 1	. Leu	Pro	Asp	Trp 5	rys	ser	ser	Leu	11e 10	Leu	мес	Ala	Tyr	11e 15	IIe	
	Ile	e Phe	Leu	Thr 20	Gly	Leu	Pro	Ala	Asn 25	Leu	Leu	Ala	Leu	Arg 30	Ala	Phe	
	Va.	l Gly	Arg 35	Ile	Arg	Gln	Pro	Gln 40	Pro	Ala	Pro	Val	His 45	Ile	Leu	Leu	
20	Let	Ser 50	Leu	Thr	Leu	Ala	Asp 55	Leu	Leu	Leu	Leu	Leu 60	Leu	Leu	Pro	Phe	
	Ly: 65	s Ile	Ile	Glu	Ala	Ala 70	Ser	Asn	Phe	Arg	Trp 75	Tyr	Leu	Pro	Lys	Val 80	
25	Va	l Cys	Ala	Leu	Thr 85	Ser	Phe	Gly	Phe	Tyr 90	Ser	Ser	Ile	Tyr	Cys 95	Ser	
	Th	r Trp	Leu	Leu 100	Ala	Gly	Ile	Ser	Ile 105	Glu	Arg	Tyr	Leu	Gly 110	Val	Ala	
	Pho	e Pro	Val 115	Gln	Tyr	Lys	Leu	Ser 120	Arg	Arg	Pro	Leu	Tyr 125	Gly	Val	Ile	
30	Ala	a Ala 130	Leu	Val	Ala	Trp	Val 135	Met	Ser	Phe	Gly	His 140	Cys	Thr	Ile	Val	
	11c	e Ile 5	Val	Gln	Tyr	Leu 150	Asn	Thr	Thr	Glu	Gln 155	Val	Arg	Ser	-	Asn 160	
35	Gl	u Ile	Thr	Cys	Tyr 165	Glu	Asn	Phe	Thr	Asp 170	Asn	Gln	Leu	Asp	Val 175	Val	

		Leu	Pro	Val	Arg 180	Leu	Glu	Leu	Cys	Leu 185		Leu	Phe	Phe	Ile 190		Met
		Ala	Val	Thr 195	Ile	Phe	Cys	Tyr	Trp 200	Arg	Phe	Val	Trp	Ile 205	Met	Leu	Ser
5		Gln	Pro 210	Leu	Val	Gly	Ala	Gln 215	Arg	Arg	Arg	Arg	Ala 220	Val	Gly	Leu	Ala
		Val 225	Val	Thr	Leu	Leu	Asn 230	Phe	Leu	Val	Cys	Phe 235	Gly	Pro	Tyr	Asn	Val 240
10		Ser	His	Leu	Val	Gly 245	Tyr	His	Gln	Arg	Lys 250	Ser	Pro	Trp	Trp	Arg 255	Ser
		Ile	Ala	Val	Val 260	Phe	Ser	Ser	Leu	Asn 265	Ala	Ser	Leu	Asp	Pro 270	Leu	Leu
		Phe	Tyr	Phe 275	Ser	Ser	Ser	Val	Val 280	Arg	Arg	Ala	Phe	Gly 285	Arg	Gly	Leu
15		Gln	Val 290	Leu	Arg	Asn	Gln	Gly 295	Ser	Ser	Leu	Leu	Gly 300	Arg	Arg	Gly	Lys
		Asp 305	Thr	Ala	Glu	Gly	Thr 310	Asn	Glu	Asp	Arg	Gly 315	Val	Gly	Gln	Gly	Glu 320
20		Gly	Met	Pro	Ser	Ser 325	Asp	Phe	Thr	Thr	Glu 330						
	(260)	INF	ORMA	MOITA	FOR	SEQ	ID	NO:2	259:								
25		(i)	(A) (B) (C)	LEN TYP STR	E CH	30 ucle DNES	base ic a S: s	pai cid ingl	.rs					·			
					OLOG				•	-							
									nomi		٠.						
	•	(xi) SE	QUEN	CE D	ESCR	IPTI	ON:	SEQ	ID N	0:25	9:					
	CCCAA	GCTT	C GG	GCAC	CATG	GAC.	ACCT	CCC									30
30	(261)	INF	ORMA	TION	FOR	SEQ	ID	NO:2	60:								
35		(i)	(A) (B) (C)	LEN TYP STR	E CH GTH: E: n ANDE	30 lucle: DNES	base ic a S: s	pai cid ingl	rs								
		(ii)) MO	LECU.	LE T	YPE:	DNA	(ge	nomi	c)							
		(xi)) SE	QUEN	CE DI	ESCR	IPTI	ON:	SEQ :	ID N	0:26):					

WO 00/22129 PCT/US99/23938

	ACAGG	ATCCA AATGCACAGC ACTGGTAAGC	30
	(262)	INFORMATION FOR SEQ ID NO:261:	
5		(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 25 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA (genomic)	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO:261:	
10	CTATA	ACTGG GTTACATGGT TTAAC	25
	(263)	INFORMATION FOR SEQ ID NO:262:	
15		(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA (genomic)	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO:262:	
	TTTGA	ATTCA CATATTAATT AGAGACATGG	30
20	(264)	INFORMATION FOR SEQ ID NO:263:	
25		 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 2724 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
		(ii) MOLECULE TYPE: DNA (genomic)	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO:263:	
	ATGGAC	PACCT CCCGGCTCGG TGTGCTCCTG TCCTTGCCTG TGCTGCTGCA GCTGGCGACC	60
	GGGGG	CAGCT CTCCCAGGTC TGGTGTGTTG CTGAGGGGCT GCCCCACACA CTGTCATTGC	120
30	GAGCC	CGACG GCAGGATGTT GCTCAGGGTG GACTGCTCCG ACCTGGGGCT CTCGGAGCTG	180
	CCTTCC	CAACC TCAGCGTCTT CAGCTCCTAC CTAGACCTCA GTATGAACAA CATCAGTCAG	240
	CTGCTC	CCGA ATCCCCTGCC CAGTCTCCGC TTCCTGGAGG AGTTACGTCT TGCGGGAAAC	300
	GCTCTG	SACAT ACATTCCCAA GGGAGCATTC ACTGGCCTTT ACAGTCTTAA AGTTCTTATG	360

CTGCAGAATA ATCAGCTAAG ACACGTACCC ACAGAAGCTC TGCAGAATTT GCGAAGCCTT 420 CAATCCCTGC GTCTGGATGC TAACCACATC AGCTATGTGC CCCCAAGCTG TTTCAGTGGC 480 CTGCATTCCC TGAGGCACCT GTGGCTGGAT GACAATGCGT TAACAGAAAT CCCCGTCCAG GCTTTTAGAA GTTTATCGGC ATTGCAAGCC ATGACCTTGG CCCTGAACAA AATACACCAC 600 ATACCAGACT ATGCCTTTGG AAACCTCTCC AGCTTGGTAG TTCTACATCT CCATAACAAT 660 AGAATCCACT CCCTGGGAAA GAAATGCTTT GATGGGCTCC ACAGCCTAGA GACTTTAGAT 720 TTAAATTACA ATAACCTTGA TGAATTCCCC ACTGCAATTA GGACACTCTC CAACCTTAAA 780 GAACTAGGAT TTCATAGCAA CAATATCAGG TCGATACCTG AGAAAGCATT TGTAGGCAAC CCTTCTCTTA TTACAATACA TTTCTATGAC AATCCCATCC AATTTGTTGG GAGATCTGCT TTTCAACATT TACCTGAACT AAGAACACTG ACTCTGAATG GTGCCTCACA AATAACTGAA TTTCCTGATT TAACTGGAAC TGCAAACCTG GAGAGTCTGA CTTTAACTGG AGCACAGATC 1020 TCATCTCTTC CTCAAACCGT CTGCAATCAG TTACCTAATC TCCAAGTGCT AGATCTGTCT 1080 TACAACCTAT TAGAAGATTT ACCCAGTTTT TCAGTCTGCC AAAAGCTTCA GAAAATTGAC 1140 CTAAGACATA ATGAAATCTA CGAAATTAAA GTTGACACTT TCCAGCAGTT GCTTAGCCTC 1200 15 CGATCGCTGA ATTTGGCTTG GAACAAATT GCTATTATTC ACCCCAATGC ATTTTCCACT 1260 TTGCCATCCC TAATAAAGCT GGACCTATCG TCCAACCTCC TGTCGTCTTT TCCTATAACT 1320 GGGTTACATG GTTTAACTCA CTTAAAATTA ACAGGAAATC ATGCCTTACA GAGCTTGATA 1380 TCATCTGAAA ACTTTCCAGA ACTCAAGGTT ATAGAAATGC CTTATGCTTA CCAGTGCTGT GCATTTGGAG TGTGTGAGAA TGCCTATAAG ATTTCTAATC AATGGAATAA AGGTGACAAC 1500 AGCAGTATGG ACGACCTTCA TAAGAAAGAT GCTGGAATGT TTCAGGCTCA AGATGAACGT 1560 GACCTTGAAG ATTTCCTGCT TGACTTTGAG GAAGACCTGA AAGCCCTTCA TTCAGTGCAG 1620 TGTTCACCTT CCCCAGGCCC CTTCAAACCC TGTGAACACC TGCTTGATGG CTGGCTGATC 1680 AGAATTGGAG TGTGGACCAT AGCAGTTCTG GCACTTACTT GTAATGCTTT GGTGACTTCA 1740 ACAGTTTTCA GATCCCCTCT GTACATTTCC CCCATTAAAC TGTTAATTGG GGTCATCGCA 1800 GCAGTGAACA TGCTCACGGG AGTCTCCAGT GCCGTGCTGG CTGGTGTGGA TGCGTTCACT 1860 TTTGGCAGCT TTGCACGACA TGGTGCCTGG TGGGAGAATG GGGTTGGTTG CCATGTCATT 1920 GGTTTTTTGT CCATTTTTGC TTCAGAATCA TCTGTTTTCC TGCTTACTCT GGCAGCCCTG 1980 GAGCGTGGGT TCTCTGTGAA ATATTCTGCA AAATTTGAAA CGAAAGCTCC ATTTTCTAGC 2040

WO 00/22129 PCT/US99/23938

217

CTGAAAGTAA TCATTTTGCT CTGTGCCCTG CTGGCCTTGA CCATGGCCGC AGTTCCCCTG 2100

	CTGGGTGG	CA G	CAAG	TATG	G CG	CCTC	CCCT	CTC	TGCC	TGC	CTTI	GCCI	TT 1	GGGG	AGCC	:C 2	2160
	AGCACCAT	GG G	CTAC	ATGGT	r cg	CTCT	CATC	TTG	CTCA	ATT	CCCI	TTGC	TT C	CTCA	TGAT	'G 2	2220
	ACCATTGC	CT A	CACC	AAGC'	CT	ACTG	CAAT	TTG	GACA	AGG	GAGA	.CCTG	GA G	AATA	TTTG	G 2	2280
5	GACTGCTC	TA T	GGTA	AAACA	A CA	TTGC	CCTG	TTG	CTCT	TCA	CCAA	.CTGC	AT C	CTAA	ACTG	C 2	2340
	CCTGTGGC	TT T	CTTG	TCCTI	CT	CCTC	TTTA	ATA	AACC	TTA	CATT	TATC	AG T	CCTG	AAGT	A 2	400
	ATTAAGTT	та т	CCTT	CTGG1	GG	TAGT	CCCA	CTT	CCTG	CAT	GTCT	CAAT	cc c	CTTC	TCTA	.C 2	460
	ATCTTGTT	CA A	TCCT	CACTI	TA	AGGA	GGAT	CTG	GTGA	gcċ	TGAG	AAAG	CA A	ACCT	ACGT	C 2	520
	TGGACAAG	AT C	AAAA	CACCO	: AA	GCTT	GATG	TCA	ATTA	ACT	CTGA	TGAT	GT C	GAAA	AACA	G 2	580
10	TCCTGTGA	CT C	AACT	CAAGO	CT	TGGT.	AACC	TTT.	ACCA	GCT	CCAG	CATC.	AC T	TATG	ACCT	G 2	640
	CCTCCCAG	TT C	CGTG	CCATC	AC	CAGC	TTAT	CCA	GTGA	CTG	AGAG	CTGC	CA T	CTTT	CCTC	T .2	700
	GTGGCATT	rg T	CCCA'	TGTCI	CT.	AA										2	724
	(265) IN	FORM	ATIO	N FOR	SE	Q ID	NO:	264:									
15		(A (B (C (D) LEI) TYI) STI	CE CH NGTH: PE: a RANDE POLOG	90 min DNE	7 am o ac SS: not:	ino a id relev	acid: vant									
20	(x :	i) S1	EQUEI	NCE D	ESC	- RIPT:	ION:	SEQ	ID 1	NO : 2	64:						
	Met 1	Asp	Thr	Ser	Arg 5	Leu	Gly	Val	Leu	Leu 10	Ser	Leu	Pro	Val	Leu 15	Leu	
	Gln	Leu	Ala	Thr 20	Gly	Gly	Ser	Ser	Pro 25	Arg	Ser	Gly	Val	Leu 30	Leu	Arg	
25	Gly	Cys	Pro 35	Thr	His	Cys	His	Cys 40	Glu	Pro	Asp	Gly	Arg 45	Met	Leu	Leu	
	Arg	Val 50	Asp	Cys	Ser	Asp	Leu 55	Gly	Leu	Ser	Glu	Leu 60	Pro	Ser	Asn	Leu	
30	Ser 65	Val	Phe	Thr	Ser	Tyr 70	Leu	Asp	Leu	Ser	Met 75	Asn	Asn	Ile	Ser	Gln 80	•
	Leu	Leu	Pro	Asn	Pro 85	Leu	Pro	Ser	Leu	Arg 90	Phe	Leu	Glu	Glu	Leu 95	Arg	
	Leu	Ala	Gly	Asn .	Ala	Leu	Thr	Tyr	Ile	Pro	Lys	Gly	Ala	Phe	Thr	Gly	

	:			100				ς.	105		٠.		-	110		
	Leu	Tyr	Ser 115	Leu	Lys	Val	Leu	Met 120		Gln	. Asn	Asn	Gln 125		Arg	His
5	Val	Pro 130	Thr	Glu	Ala	Leu	Gln 135		Leu	Arg	Ser	Leu 140		Ser	Leu	Arg
	Leu 145	Asp	Ala	Asn	His	Ile 150	Ser	Tyr	Val	Pro	Pro 155	Ser	Суѕ	Phe	Ser	Gly 160
	Leu	His	Ser	Leu	Arg 165	His	Leu	Trp	Leu	Asp 170		Asn	Ala	Leu	Thr 175	Glu
10	Ile	Pro	Val	Gln 180	Ala	Phe	Arg	Ser	Leu 185	Ser	Ala	Leu	Gln	Ala 190	Met	Thr
	Leu	Ala	Leu 195	Asn	Lys	Ile	His	His 200	Ile	Pro	Asp		Ala 205		Gly	Asn
15	Leu	Ser 210	Ser	Leu	Val	Val	Leu 215	His	Leu	His	Asn	Asn 220	Arg	Ile	His	Ser
	Leu 225	Gly	Lys	Lys	Cys	Phe 230	Asp	Gly	Leu	His	Ser 235	Leu	Glu	Thr	Leu	Asp 240
	Leu	Asn	Tyr	Asn	Asn 245	Leu	Asp	Glu	Phe	Pro 250		Ala	Ile	Arg	Thr 255	Leu
20	Ser	Asn	Leu	Lys 260	Glu	Leu	Gly	Phe	His 265	Ser	Asn	Asn	Ile	Arg 270	Ser	Ile
	Pro	Glu	Lys 275	Ala	Phe	Val	Gly	Asn 280	Pro	Ser	Leu		Thr 285		His	Phe
25	Tyr	Asp 290	Asn	Pro	Ile	Gln	Phe 295		Gly	Arg	Ser	Ala 300	Phe	Gln	His	Leu
	Pro 305	Glu	Leu	Arg	Thr	Leu 310	Thr	Leu	Asn	Gly	Ala 315	Ser	Gln	Ile	Thr	Glu 320
	Phe	Pro	Asp		Thr 325	Gly	Thr	Ala	Asn	Leu 330	Glu	Ser	Leu	Thr	Leu 335	Thr
30	Gly	Ala	Gln	Ile 340	Ser	Ser	Leu	Pro	Gln 345	Thr	Val	Cys		Gln 350	Leu	Pro
	Asn	Leu	Gln 355	Val	Leu	Asp	Leu	Ser 360	Tyr	Asn	Leu		Glu 365	Asp	Leu	Pro
35	Ser	Phe 370	Ser	Val	Cys		Lys 375	Leu	Gln	Lys	Ile	Asp 380	Leu	Arg	His	Asn
	Glu 385	Ile	Tyr	Glu	Ile	Lys 390	Val	Asp	Thr	Phe	Gln 395	Gln	Leu	Leu	Ser	Leu 400

		Arg	Ser	Leu	Asn ,	Leu 405	Ala	Trp	Asn	Lys	Ile 410		Ile	Ile	His	Pro 415	
	7	Ala	Phe	Ser	Thr 420	Leu	Pro	Ser	Leu	1le 425	Lys	Leu	Asp	Leu	Ser 430		Asr
5	I	Leu	Leu	Ser 435	Ser	Phe	Pro	Ile	Thr 440	Gly	Leu	His	Gly	Leu 445		His	Leu
	I	ys	Leu 450	Thr	Gly	Asn	His	Ala 455		Gln	Ser	Leu	Ile 460	Ser	Ser	Glu	Asn
10		he 65	Pro	Glu	Leu	Lys	Val 470	Ile	Glu	Met	Pro	Tyr 475		Tyr	Gln	Суз	Cys 480
	Æ	la	Phe	Gly	Val	Cys 485	Glu	Asn	Ala	Tyr	Lys 490	Ile	Ser	Asn	Gln	Trp 495	Asn
,	L	ys	Gly	Asp	Asn 500	Ser	Ser	Met	Asp	Asp 505	Leu	His	Lys	Lys	Asp 510	Ala	Gly
15	M	let	Phe	Gln 515	Ala	Gln	Asp	Glu	Arg 520	Asp	Leu	Glu	Asp	Phe 525	Leu	Leu	Asp
	P		Glu 530	Glu	Asp	Leu	Lys	Ala 535	Leu	His	Ser	Val	Gln 540	Cys	Ser	Pro	Ser
20		ro 45	Gly	Pro	Phe	Lys	Pro 550	Cys	Glu	His	Leu	Leu 555	Asp	Gly	Trp	Leu	Ile 560
	A	rg	Ile	Gly	Val	Trp 565	Thr	Ile	Ala	Val	Leu 570	·Ala	Leu	Thr	Cys	Asn 575	Ala
	L	eu '	Val	Thr	Ser 580	Thr	Val	Phe	Arg	Ser 585	Pro	Leu	Tyr	Ile	Ser 590	Pro	Ile
25	L	ys :	Leu	Leu 595	Ile	Gly	Val	Ile	Ala 600	Ala	Val	Asn	Met	Leu 605	Thr	Gly	Val
	s		Ser 610	Ala	Val	Leu	Ala	Gly 615	Val	Asp	Ala	Phe	Thr 620	Phe	Gly	Ser	Phe
30		la 2 25	Arg	His	Gly	Ala	Trp 630	Trp	Glu	Asn	Gly	Val 635	Gly	Cys	His	Val	Ile 640
	G	ly i	Phe	Leu	Ser	Ile 645	Phe	Ala	Ser	Glu	Ser 650	Ser	Val	Phe	Leu	Leu 655	Thr
	L	eu i	Ala		Leu 660	Glu	Arg	Gly	Phe	Ser 665	Val	Lys	Tyr	Ser	Ala 670	Lys	Phe
35	G:	lu :		Lys 675	Ala	Pro	Phe	Ser	Ser 680	Leu	Lys	Val	Ile	Ile 685	Leu	Leu	Cys
	A.	la 1	Leu	Leu	Ala	Leu	Thr	Met	Ala	Ala	Val	Pro	Leu	Leu	Gly	Gly	Ser

		690					695		*	• •	٠.	700				
	Lys 705	Tyr	Gly	Ala	Ser	Pro 710	Leu	Cys	Leu	Pro	Leu 715	Pro	Phe	Gly	Glu	Pro 720
5	Ser	Thr	Met	Gly	Tyr 725	Met	Val	Ala	Leu	Ile 730	Leu	Leu	Asn	Ser	Leu 735	Cys
	Phe	Leu	Met	Met 740	Thr	Ile	Ala	Tyr	Thr 745	Lys	Leu	Tyr	Суз	Asn 750	Leu	Asp
	Lys	Gly	Asp 755	Leu	Glu	Asn	Ile	Trp 760	Asp	Cys	Ser	Met	Val 765	Lys	His	Ile
10	Ala	Leu 770	Leu	Leu	Phe	Thr	Asn 775	Cys	Ile	Leu	Asn	Cys 780	Pro	Val	Ala	Phe
	Leu 785	Ser	Phe	Ser	Ser	Leu 790	Ile	Asn	Leu	Thr	Phe 795	Ile	Ser	Pro	Glu	Val 800
15	Ile	Lys	Phe	Ile	Leu 805	Leu	Val	Val	Val	Pro 810	Leu	Pro	Ala	Cys	Leu 815	Asn
	Pro	Leu	Leu	Tyr 820	Ile	Leu	Phe	Asn	Pro 825	His	Phe	Lys	Glu	Asp 830	Leu	Val
	Ser		Arg 835	Lys	Gln	Thr	Tyr	Val 840	_	Thr	Arg	Ser	Lys 845	His	Pro	Ser
20	Leu	Met 850	Ser	Ile	Asn	Ser	Asp 855	Asp	Val	Glu	Lys	Gln 860	Ser	Cys	Asp	Ser
	Thr 865	Gln	Ala	Leu	Val	Thr 870	Phe	Thr	Ser	Ser	Ser 875	Ile	Thr	Tyr	Asp	Leu 880
25	Pro	Pro	Ser	Ser	Val 885	Pro	Ser	Pro	Ala	Tyr 890	Pro	Val	Thr	Glu	Ser 895	Cys
	His	Leu	Ser	Ser 900	Val	Ala	Phe		Pro 905	Cys	Leu	·				
	(266) IN	FORM	ATIO	N FO	R SE	Q ID	NO:	265:								
30	(i	(A (B (C	QUENC) LE) TY) ST) TO	NGTH PE: 1 RAND	: 30 nucl EDNE	base eic a	e pa: acid sing:	irs								
	(i	i) M	OLEC	ULE '	TYPE	: DN	A (g	enom	ic)							
35	(x	i) S	EQUE	NCE 1	DESC	RIPT	ION:	SEQ	ID :	NO:2	65:	٠.				

CGGAAGCTGC GGGCCAAATG GGTGGCCGGC

	(267) INFORMATION FOR SEQ ID NO:266:	
5	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 27 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic) (xi) SEQUENCE DESCRIPTION: SEQ ID NO:266:	
	CAGAGGAGGG TGAAGGGGCT GTTGGCG	27
10	(268) INFORMATION FOR SEQ ID NO:267:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	:
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:267:	
	GGCGGCGCCG AGCCAAGGGG CTGGCTGTGG	30
	(269) INFORMATION FOR SEQ ID NO:268:	
20	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 32 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	4
25	(ii) MOLECULE TYPE: DNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:268:	•
	GGGACTGCTC TATGAAAAAA CACATTGCCC TG	32
	(270) INFORMATION FOR SEQ ID NO:269:	
30	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1071 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA (genomic)	
35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:269:	
	ATGAATGGGG TCTCGGAGGG GACCAGAGGC TGCAGTGACA GGCAACCTGG GGTCCTGACA	60

	CGTGATCGCT	CTTGTTCCAG	GAAGATGAAC	TCTTCCGGAT	GCCTGTCTGA	GGAGGTGGGG	120
	TCCCTCCGCC	CACTGACTGT	GGTTATCCTG	TCTGCGTCCA	TTGTCGTCGG	AGTGCTGGGC	180
4	AATGGGCTGG	TGCTGTGGAT	GACTGTCTTC	CGTATGGCAC	GCACGGTCTC	CACCGTCTGC	240
	TTCTTCCACC	TGGCCCTTGC	CGATTTCATG	CTCTCACTGT	CTCTGCCCAT	TGCCATGTAC	300
5	TATATTGTCT	CCAGGCAGTG	GCTCCTCGGA	GAGTGGGCCT	GCAAACTCTA	CATCACCTTT	360
	GTGTTCCTCA	GCTACTTTGC	CAGTAACTGC	CTCCTTGTCT	TCATCTCTGT	GGACCGTTGC	420
	ATCTCTGTCC	TCTACCCCGT	CTGGGCCCTG	AACCACCGCA	CTGTGCAGCG	GGCGAGCTGG	480
	CTGGCCTTTG	GGGTGTGGCT	CCTGGCCGCC	GCCTTGTGCT	CTGCGCACCT	GAAATTCCGG	540
	ACAACCAGAA	AATGGAATGG	CTGTACGCAC	TGCTACTTGG	CGTTCAACTC	TGACAATGAG	600
10	ACTGCCCAGA	TTTGGATTGA	AGGGGTCGTG	GAGGGACACA	TTATAGGGAC	CATTGGCCAC	660
	TTCCTGCTGG	GCTTCCTGGG	GCCCTTAGCA	ATCATAGGCA	CCTGCGCCCA	CCTCATCCGG	720
•	GCCAAGCTCT	TGCGGGAGGG	CTGGGTCCAT	GCCAACCGGC	CCAAGAGGCT	GCTGCTGGTG	780
	CTGGTGAGCG	CTTTCTTTAT	CTTCTGGTCC	CCGTTTAACG	TGGTGCTGTT	.GGTCCATCTG .	840
	TGGCGACGGG	TGATGCTCAA	GGAAATCTAC	CACCCCGGA	TGCTGCTCAT	CCTCCAGGCT	900
15	AGCTTTGCCT	TGGGCTGTGT	CAÀCAGCAGC	CTCAACCCCT	TCCTCTACGT	CTTCGTTGGC	960
	AGAGATTTCC	AAGAAAAGTT	TTTCCAGTCT	TTGACTTCTG.	CCCTGGCGAG	GGCGTTTGGA	1020
	GAGGAGGAGT	TTCTGTCATC	CTGTCCCCGT	GGCAACGCCC	CCCGGGAATG	A	1071
	(271) INFOR	RMATION FOR	SEQ ID NO:2	270:		•	

(i) SEQUENCE CHARACTERISTICS:

20

- (A) LENGTH: 356 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:270:

Met Asn Gly Val Ser Glu Gly Thr Arg Gly Cys Ser Asp Arg Gln Pro 1 5 10 15

Gly Val Leu Thr Arg Asp Arg Ser Cys Ser Arg Lys Met Asn Ser Ser 20 25 30

30 Gly Cys Leu Ser Glu Glu Val Gly Ser Leu Arg Pro Leu Thr Val Val 35 40 45

	Ile	Leu 50	Ser	Ala	Ser	lle	Val	Val	Gly	Val	Leu	Gly 60	Asn	Gly	Leu	Val
	Leu 65	Trp	Met	Thr	Val	Phe 70	Arg	Met	Ala	Arg	Thr 75	Val	Ser	Thr	Val	Cys 80
5	Phe	Phe	His	Leu	Ala 85	Leu	Ala	Asp	Phe	Met 90	Leu	Ser	Leu	Ser	Leu 95	Pro
	Ile	Ala	Met	Tyr 100	Tyr	Ile	Val	Ser	Arg 105	Gln	Trp	Leu	Leu	Gly 110	Glu	Trp
10	Ala	Cys	Lys 115	Leu	Tyr	Ile	Thr	Phe 120	Val	Phe	Leu	Ser	Tyr 125	Phe	Ala	Ser
	Asn	Cys 130	Leu	Leu	Val	Phe	Ile 135	Ser	Val	Asp	Arg	Cys 140	Ile	Ser	Val	Leu
÷.	Tyr 145	Pro	Val	Trp	Ala	Leu 150	Asn	His	Arg	Thr	Val 155	Gln	Arg	Ala	Ser	Trp 160
15 .	Leu	Ala	Phe	Gly	Val 165	Trp	Leu	Leu	Ala	Ala 170	Ala	Leu	Cys	Ser	Ala 175	<u>Ḥis</u>
ż	Leu	Lys	Phe	Arg 180	Thr	Thr	Arg	Lys	Trp 185	Asn	Gly	Cys	Thr	His 190	Cys	Tyr
20	Leu	Ala	Phe 195	Asn	Ser	Asp	Asn	Glu 200	Thr	Ala	Gln	Ile	Trp 205	Ile	Glu	Gly
	Val	Val 210	Glu	Gly	His	Ile	Ile 215	Gly	Thr	Ile	Gly	His 220	Phe	Leu	Leu	Gly
	Phe 225	Leu	Gly	Pro	Leu	Ala 230	Ile	Ile	Gly	Thr	Cys 235	Ala	His	Leu	Ile	Arg 240
25	Ala	Lys	Leu	Leu	Arg 245	Glu	Gly	Trp	Val	His 250	Ala	Asn	Arg	Pro	Lys 255	Arg
	Leu	Leu	Leu	Val 260	Leu	Val	Ser	Ala	Phe 265	Phe	Ile	Phe	Trp	Ser 270	Pro	Phe
30	Asn	Val	Val 275	Leu	Leu	Val	His	Leu 280	Trp	Arg	Arg	Val	Met 285	Leu	Lys	Glu
	Ile	Tyr 290	His	Pro	Arg	Met	Leu 295	Leu	Ile	Leu	Gln	Ala 300	Ser	Phe	Ala	Leu
	Gly 305	Cys	Val	Asn	Ser	Ser 310	Leu	Asn	Pro	Phe	Leu 315	Tyr	Val	Phe	Val	Gly 320
35	Arg	Asp	Phe	Gln	Glu 325	Lys	Phe	Phe	Gln	Ser 330	Leu	Thr	Ser	Ala	Leu 335	Ala
	Arg	Ala	Phe	Gly	Glu	Glu	Glu	Phe	Leu	Ser	Ser	Cys	Pro	Ara	Glv	Asn

	340	345	•	350
--	-----	-----	---	-----

Ala Pro Arg Glu 355

(272) INFORMATION	FOR	SEQ	ID	NO:271:
-------------------	-----	-----	----	---------

5	(i)	SEQ	UENCE	CH	ARAC'	TERIS:	rics:
		(A)	LENG	TH:	903	base	pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

10

- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:271:

ATGGACCTGC CCCCGCAGCT CTCCTTCGGC CTCTATGTGG CCGCCTTTGC GCTGGGCTTC 60 CCGCTCAACG TCCTGGCCAT CCGAGGCGCG ACGGCCCACG CCCGGCTCCG TCTCACCCCT 120 15 AGCCTGGTCT ACGCCCTGAA CCTGGGCTGC TCCGACCTGC TGCTGACAGT CTCTCTGCCC 180 CTGAAGGCGG TGGAGGCGCT AGCCTCCGGG GCCTGGCCTC TGCCGGCCTC GCTGTGCCCC 240 GTCTTCGCGG TGGCCCACTT CTTCCCACTC TATGCCGGCG GGGGCTTCCT GGCCGCCCTG 300 AGTGCAGGCC GCTACCTGGG AGCAGCCTTC CCCTTGGGCT ACCAAGCCTT CCGGAGGCCG 360 TGCTATTCCT GGGGGGTGTG CGCGGCCATC TGGGCCCTCG TCCTGTGTCA CCTGGGTCTG 420 20 GTCTTTGGGT TGGAGGCTCC AGGAGGCTGG CTGGACCACA GCAACACCTC CCTGGGCATC 480 AACACACCGG TCAACGGCTC TCCGGTCTGC CTGGAGGCCT GGGACCCGGC CTCTGCCGGC 540 CCGGCCCGCT TCAGCCTCTC TCTCCTGCTC TTTTTTCTGC CCTTGGCCAT CACAGCCTTC 600 TGCTACGTGG GCTGCCTCCG GGCACTGGCC CGCTCCGGCC TGACGCACAG GCGGAAGCTG 660 CGGGCCAAAT GGGTGGCCGG CGGGGCCCTC CTCACGCTGC TGCTCTGCGT AGGACCCTAC 720 25 AACGCCTCCA ACGTGGCCAG CTTCCTGTAC CCCAATCTAG GAGGCTCCTG GCGGAAGCTG 780 GGGCTCATCA CGGGTGCCTG GAGTGTGGTG CTTAATCCGC TGGTGACCGG TTACTTGGGA 840 AGGGGTCCTG GCCTGAAGAC AGTGTGTGCG GCAAGAACGC AAGGGGGCAA GTCCCAGAAG 900 TAA 903

(273) INFORMATION FOR SEQ ID NO:272:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 300 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:

- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:272:

	(xi	L) SI	EQUEN	ICE I	DESCI	RIPTI	ON:	SEQ	ID 1	10:27	72:					
5	Met 1	Asp	Leu	Pro	Pro 5	Gln	Leu	Ser	Phe ·	Gly 10	Leu	Tyr	Val	Ala	Ala 15	Phe
	Ala	Leu	Gly	Phe 20	Pro	Leu	Asn	Val	Leu 25	Ala	Ile	Arg	Gly	Ala 30	Thr	Ala
	His	Ala	Arg 35	Leu	Arg	Leu	Thr	Pro 40	Ser	Leu	Val	Tyr	Ala 45	Leu	Asn	Leu
10	Gly	Cys 50	Ser	Asp	Leu	Leu	Leu 55	Thr	Val	Ser	Leu	Pro 60	Leu	Lys	Ala	Val
	Glu 65	Ala	Leu	Ala	Ser	Gly 70	Ala	Trp	Pro	Leu	Pro 75	Ala	Ser	Leu	Cys	Pro 80
15	Val	Phe	Ala	Val	Ala 85	His	Phe	Phe	Pro	Leu 90	Tyr	Ala	Gly	Gly	Gly 95	Phe
	Leu	Ala	Ala	Leu 100	Ser	Ala	Gly	Arg	Tyr 105	Leu	Gly	Ala	Ala	Phe 110	Pro	Leu
	Gly	Tyr	Gln 115	Ala	Phe.	Arg	Arg	Pro 120	Cys	Tyr	Ser	Trp	Gly 125	Val	Cys	Ala
20	Ala	Ile 130	Trp	Ala	Leu	Val	Leu 135	Cys	His	Leu	Gly	Leu 140	Val	Phe	Gly	Leu
	Glu 145	Ala	Pro	Gly	Gly	Trp 150	Leu	Asp	His	Ser	Asn 155	Thr	Ser	Leu	Gly	Ile 160
25	Asn	Thr	Pro	Val	Asn 165	Gly	Ser	Pro	Val	Cys 170	Leu	Glu	Ala	Trp	Asp 175	Pro
	Ala	Ser	Ala	Gly 180	Pro	Ala	Arg	Phe	Ser 185	Leu	Ser	Leu	Leu	Leu 190	Phe	Phe
	Leu	Pro	Leu 195	Ala	Ile	Thr	Ala	Phe 200	Cys	Tyr	Val	Gly	Cys 205	Leu	Arg	Ala
30	Leu	Ala 210	Arg	Ser	Gly	Leu	Thr 215	His	Arg	Arg	Lys	Leu 220	Arg	Ala	Lys	Trp
	Val 225	Ala	Gly	Gly	Ala	Leu 230	Leu	Thr	Leu	Leu	Leu 235	Cys	Val	Gly	Pro	Tyr 240
35	Asn	Ala	Ser	Asn	Val 245	Ala	Ser	Phe	Leu	Tyr 250	Pro	Asn	Leu	Gly	Gly 255	Ser
	Trp	Arg	Lys	Leu	Gly	Leu	Ile	Thr	Gly	Ala	Trp	Ser	Val	Val	Leu	Asn

260		265	 . 270

Pro Leu Val Thr Gly Tyr Leu Gly Arg Gly Pro Gly Leu Lys Thr Val 275 280 285

Cys Ala Ala Arg Thr Gln Gly Gly Lys Ser Gln Lys
5 290 295 300

(274) INFORMATION FOR SEQ ID NO:273:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1041 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:273:

			•			•	
•	ATGGATACAG	GCCCCGACCA	GTCCTACTTC	TCCGGCAATC	ACTGGTTCGT	CTTCTCGGTG	60
15	TACCTTCTCA	CTTTCCTGGT	GGGGCTCCCC	CTCAACCTGC	TGGCCCTGGT	GGTCTTCGTG	120
	GGCAAGCTGC	AGCGCCGCCC	GGTGGCCGTG	GACGTGCTCC	TGCTCAACCT	GACCGCCTCG	180
	GACCTGCTCC	TGCTGCTGTT	CCTGCCTTTC	CGCATGGTGG	AGGCAGCCAA	TGGCATGCAC	240
	TGGCCCCTGC	CCTTCATCCT	CTGCCCACTC	TCTGGATTCA	TCTTCTTCAC	CACCATCTAT	300
	CTCACCGCCC	TCTTCCTGGC	AGCTGTGAGC	ATTGAACGCT	TCCTGAGTGT	GGCCCACCCA	360
20	CTGTGGTACA	AGACCCGGCC	GAGGCTGGG	CAGGCAGGTC	TGGTGAGTGT	GGCCTGCTGG	420
	CTGTTGGCCT	CTGCTCACTG	CAGCGTGGTC	TACGTCATAG	AATTCTCAGG	GGACATCTCC	480
	CACAGCCAGG	GCACCAATGG	GACCTGCTAC	CTGGAGTTCC	GGAAGGACCA	GCTAGCCATC	540
	CTCCTGCCCG	TGCGGCTGGA	GATGGCTGTG	GTCCTCTTTG	TGGTCCCGCT	GATCATCACC	600
	AGCTACTGCT	ACAGCCGCCT	GGTGTGGATC	CTCGGCAGAG	GGGGCAGCCA	CCGCCGGCAG	660
25	AGGAGGGTGA	AGGGGCTGTT	GGCGGCCACG	CTGCTCAACT	TCCTTGTCTG	CTTTGGGCCC	720
٠	TACAACGTGT	CCCATGTCGT	GGGCTATATC	TGCGGTGAAA	GCCCGGCATG	GAGGATCTAC	780
	GTGACGCTTC	TCAGCACCCT	GAACTCCTGT	GTCGACCCCT	TTGTCTACTA	CTTCTCCTCC	840
	TCCGGGTTCC	AAGCCGACTT	TCATGAGCTG	CTGAGGAGGT	TGTGTGGGCT	CTGGGGCCAG	900
	TGGCAGCAGG	AGAGCAGCAT	GGAGCTGAAG	GAGCAGAAGG	GAGGGGAGGA	GCAGAGAGCG	960
30	GACCGACCAG	CTGAAAGAAA	GACCAGTGAA	CACTCACAGG	GCTGTGGAAC	TGGTGGCCAG	1020
	GTGGCCTGTG	CTGAAAGCTA	G				1041

(275) INFORMATION FOR SEQ ID NO:274:

35

5	. (i	(A (B	QUEN LE TY ST TO	ngth PE : RAND	: 34 amin EDNE	6 am o ac SS:	ino id	acid				-				
	(i	.i) M	OLEC	ULE '	TYPE	: pr	otei:	n								
	(x	i) S	EQUE	NCE :	DESC	RIPT	ION:	SEQ	ID	NO:2	74:					
10	Met 1	Asp	Thr	Gly	Pro 5	Asp	Gln	Ser	Tyr	Phe 10	Ser	Gly	Asn	His	Trp 15	Phe
	Val	Phe	Ser	Val 20	Tyr	Leu	Leu	Thr	Phe 25	Leu	Val	Gly	Leu	Pro 30	Leu	Asn
	Leu	Leu	Ala 35	Leu	Val	Val	Phe	Val 40	Gly	Lys	Leu	Gln	Arg 45	Arg	Pro	Val
15	Ala	Val 50	Asp	Val	Leu	Leu	Leu 55	Asn	Leu	Thr	Ala	Ser 60	Asp	Leu	Leu	Leu
	Leu 65	Leu	Phe	Leu	Pro	Phe 70	Arg	Met	Val	Glu	Ala 75	Ala	Asn	Gly	Met	His 80
20	Trp	Pro	Leu	Pro	Phe 85	Ile	Leu	Cys	Pro	Leu 90	Ser	Gly	Phe	Ile	Phe 95	Phe
	Thr	Thr	Ile	Tyr 100		Thr	Ala	Leu	Phe 105	Leu	Ala	Ala	Val	Ser 110	Ile	Glu
	Arg	Phe	Leu 115	Ser	Val	Ala	His	Pro 120	Leu	Trp	Tyr	Lys	Thr 125	Arg	Pro	Arg
25	Leu	Gly 130	Gln	Ala	Gly	Leu	Val 135	Ser	Val	Ala	Cys	Trp 140	Leu	Leu	Ala	Ser
	Ala 145	His	Cys	Ser	Val	Val 150	Tyr	Val	Ile	Glu	Phe 155	Ser	Gly	Asp	Ile	Ser 160
30	His	Ser	Gln	Gly	Thr 165	Asn	Gly	Thr	Cys	Tyr 170	Leu	Glu	Phe	Arg	Lys 175	Asp
	Gln	Leu	Ala	Ile 180	Leu	Leu	Pro	Val	Arg 185	Leu	Glu	Met	Ala	Val 190	Val	Leu
	Phe	Val	Val 195	Pro	Leu	Ile	Ile	Thr 200	Ser	Tyr	Cys	Tyr	Ser 205	Arg	Leu	Val

Trp Ile Leu Gly Arg Gly Gly Ser His Arg Arg Gln Arg Arg Val Lys

Gly Leu Leu Ala Ala Thr Leu Leu Asn Phe Leu Val Cys Phe Gly Pro

	225			•		230	. •				235	, .		٠		240	
	Tyr	Asn	Val	Ser	His 245	Val	Val	Gly	Tyr	Ile 250	Cys	Gly	Glu	Ser	Pro 255	Ala	
5	Trp	Arg	Ile	Tyr 260	Val	Thr	Leu	Leu	Ser 265		Leu	Asn	Ser	Cys 270	Val	Asp	
	Pro	Phe	Val 275	Tyr	Tyr	Phe	Ser	Ser 280	Ser	Gly	Phe	Gln	Ala 285	Asp	Phe	His	
	Glu	Leu 290	Leu	Arg	Arg	Leu	Cys 295	Gly	Leu	Trp	Gly	Gln 300	Trp	Gln	Gln	Glu	
10	Ser 305	Ser	Met	Glu	Leu	Lys 310	Glu	Gln	Lys	Gly	Gly 315	Glu	Glu	Gln	Arg	Ala 320	
•	Asp	Arg	Pro	Ala	Glu 325	Arg	Lys	Thr		Glu 330	His	Ser	Gln	Gly	Cys 335	Gly	
15	Thr	Gly	Gly	Gln 340	Val	Ala	Cys	Ala	Glu 345	Ser							
	(276) IN	FORM	ATION	I FOR	SEÇ] ID	NO:2	75:									
	(i) SEQ	OUENC	CE CH	IARAC	TERI	STIC	's ·									
	, -	(A)	LEN	IGTH :	993	bas	se pa										
20		(C)	TYP STF TOP	RANDE	DNES	S: s	ingl	e									
	(i:	i) MC	DLECU	JLE T	YPE:	DNA	ı (ge	nomi	.c)				,				
	(x:	i) SE	EQUEN	ICE D	ESCR	IPTI	ON:	SEQ	ID N	10:27	5:						
•	ATGCTGCC	GG AC	TGGA	AGAG	CTC	CTTG	ATC	CTCA	TGGC	TT A	CATC	ATCA	T CI	TCCT	CACT		60
25	GGCCTCCC	rg co	CAACC	TCCT	GGC	CCTG	CGG	GCCT	TTGT	GG G	GCGG	ATCC	G CC	AGCC	CCAG	1	.20
•	CCTGCACC	rg To	CACA	TCCT	CCT	GCTG	AGC	CTGA	.CGCT	'GG C	CGAC	CTCC	T CC	TGCT	GCTG	1	80
	CTGCTGCC	T TC	'AAGA	TCAT	' CGA	.GGCT	GCG	TCGA	ACTI	CC G	CTGG	TACC	T GC	CCAA	GGTC	2	40
	GTCTGCGC	CC TC	:ACGA	GTTT	TGG	CTTC	TAC	AGCA	GCAT	CT A	.CTGC	AGCA	C GT	GGCT	CCTG	3	00
	GCGGGCAT	CA GO	ATCG	AGCG	CTA	CCTG	GGA	GTGG	CTTT	cc c	CGTG	CAGT	A CA	AGCT	CTCC	3	60
30	CGCCGGCC	rc Te	TATG	GAGT	GAT	TGCA	GCT	CTGG	TGGC	CT G	GGTT	ATGT	с ст	TTGG	TCAC	4	20
	TGCACCATO	CG TG	ATCA	TCGT	TCA	ATAC	TTG	AACA	CGAC	TG A	GCAG	GTCA	g aa	GTGG	CAAT	4	80
	GAAATTAC	CT GC	TACG	AGAA	CTT	CACC	GAT	AACC	AGTT	GG A	CGTG	GTGC	T GC	CCGT	GCGG	5	40
	CTGGAGCT	T GC	CTGG	TGCT	CTT	CTTC	ATC	CCCA	TGGC	AG T	CACC	ATCT	т ст	GCTA	CTGG	6	00

BRIGDOOID, JAKO MOOHOOK IA.

	CGTTTTGTGT GGATCATGCT CTCCCAGCCC CTTGTGGGGG CCCAGAGGCG GCGCCGAGCC	660
	AAGGGGCTGG CTGTGGTGAC GCTGCTCAAT TTCCTGGTGT GCTTCGGACC TTACAACGTG	720
	TCCCACCTGG TGGGGTATCA CCAGAGAAAA AGCCCCTGGT GGCGGTCAAT AGCCGTGGTG	780
	TTCAGTTCAC TCAACGCCAG TCTGGACCCC CTGCTCTTCT ATTTCTCTTC TTCAGTGGTG	840
5	CGCAGGGCAT TTGGGAGAGG GCTGCAGGTG CTGCGGAATC AGGGCTCCTC CCTGTTGGGA	900
	CGCAGAGGCA AAGACACAGC AGAGGGGACA AATGAGGACA GGGGTGTGGG TCAAGGAGAA	960
	GGGATGCCAA GTTCGGACTT CACTACAGAG TAG	993
	(277) INFORMATION FOR SEQ ID NO:276:	
10	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 330 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: (D) TOPOLOGY: not relevant (ii) MOLECULE TYPE: protein 	
15	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:276:	
	Met Leu Pro Asp Trp Lys Ser Ser Leu Ile Leu Met Ala Tyr Ile Ile 1 5 10 15	
	Ile Phe Leu Thr Gly Leu Pro Ala Asn Leu Leu Ala Leu Arg Ala Phe 20 25 30	
20	Val Gly Arg Ile Arg Gln Pro Gln Pro Ala Pro Val His Ile Leu Leu 35 40 45	
	Leu Ser Leu Thr Leu Ala Asp Leu Leu Leu Leu Leu Leu Pro Phe 50 55 60	
25	Lys Ile Ile Glu Ala Ala Ser Asn Phe Arg Trp Tyr Leu Pro Lys Val 65 70 75 80	
	Val Cys Ala Leu Thr Ser Phe Gly Phe Tyr Ser Ser Ile Tyr Cys Ser 85 90 95	
	Thr Trp Leu Leu Ala Gly Ile Ser Ile Glu Arg Tyr Leu Gly Val Ala 100 105 110	
30	Phe Pro Val Gln Tyr Lys Leu Ser Arg Arg Pro Leu Tyr Gly Val Ile 115 120 125	
	Ala Ala Leu Val Ala Trp Val Met Ser Phe Gly His Cys Thr Ile Val 130 135 140	
35	Ile Ile Val Gln Tyr Leu Asn Thr Thr Glu Gln Val Arg Ser Gly Asn 145 150 155 160	

	G)	u Il	.e. Th:	r Cys	Tyr 165	Glu	Asn	Phe	Thr	170	Asn	Gln	Leu	Asp	Val 175	
	Le	u Pr	o Val	l Arg	Leu	Glu	Leu	Cys	Leu 185	Val	Leu	Phe	Phe	Ile 190		Met
5	Al	a Va	l Thi 199	Ile	Phe	Cys	Tyr	Trp 200	Arg	Phe	Val	Trp	Ile 205	Met	Leu	Ser
	Gl	n Pr 21	o Leu 0	ı Val	Gly	Ala	Gln 215	Arg	Arg	Arg	Arg	Ala 220	Lys	Gly	Leu	Ala
10	Va 22	l Vai	l Thr	Leu	Leu	Asn 230	Phe	Leu	Val	Cys	Phe 235	Gly	Pro	Tyr	Asn	Val 240
	Se	r Hi	s Leu	Val	Gly 245	Tyr	His	Gln	Arg	Lys 250	Ser	Pro	Trp	Trp	Arg 255	Ser
	Il	≥ Ala	a Val	Val 260	Phe	Ser	Ser	Leu	Asn 265	Ala	Ser	Leu	Asp	Pro 270	Leu	Leu
15	Phe	е Туг	275	Ser	Ser	Ser	Val	Val 280	Arg	Arg	Ala		Gly 285	Arg	Gly	Leu
		230	,	Arg			295					300				
20	Asg 305	Thr	Ala	Glu	Gly	Thr 310	Asn	Glu	Asp		Gly ' 315	Val	Gly	Gln		Glu 320
	Gly	Met	Pro	Ser	Ser 325	Asp	Phe	Thr		Glu 330						
	(278) IN	FORM	ATION	N FOR	SEQ	ID 1	NO:2	77:								
25	(i	(A (B (C) LEN) TYE) STR	CE CH NGTH: PE: n NANDE	272 ucle: DNES:	4 bas ic ac S: s:	se pa cid ingle	airs								
	(i			LE T												
30																
				CE DI												
	ATGGACAC							*								60
	GGGGGCAG															120
	GAGCCCGA															180
	CCTTCCAA															
35	CTGCTCCCC	IA A	CCCC	TGCC	CAGI	CTCC	GC T	TCCT	'GGAG	G AG	TTAC	GTCT	TGC	GGGA	AAC	300

WO 00/22129 PCT/US99/23938

•	GCTCTGACAT	ACATTCCCAA	GGGAGCATTC	ACTGGCCTTT	ACAGTCTTAA	AGTTCTTATG	360
	CTGCAGAATA	ATCAGCTAAG	ACACGTACCC	ACAGAAGCTC	TGCAGAATTI	GCGAAGCCTT	420
	CAATCCCTGC	GTCTGGATGC	TAACCACATC	AGCTATGTGC	CCCCAAGCTG	TTTCAGTGGC	400
	CTGCATTCCC	TGAGGCACCT	GTGGCTGGAT	GACAATGCGT	TAACAGAAAT	CCCCGTCCAG	540
5	GCTTTTAGAA	GTTTATCGGC	ATTGCAAGCC	ATGACCTTGG	CCCTGAACAA	AATACACCAC	600
	ATACCAGACT	ATGCCTTTGG	AAACCTCTCC	AGCTTGGTAG	TTCTACATCT	CCATAACAAT	660
	AGAATCCACT	CCCTGGGAAA	GAAATGCTTT	GATGGGCTCC	ACAGCCTAGA	GACTTTAGAT	720
	TTAAATTACA	ATAACCTTGA	TGAATTCCCC	ACTGCAATTA	GGACACTCTC	CAACCTTAAA	780
	GAACTAGGAT	TTCATAGCAA	CAATATCAGG	TCGATACCTG	AGAAAGCATT	TGTAGGCAAC	840
10	CCTTCTCTTA	TTACAATACA	TTTCTATGAC	AATCCCATCC	AATTTGTTGG	GAGATCTGCT	900
	TTTCAACATT	TACCTGAACT	AAGAACACTG	ACTCTGAATG	GTGCCTCACA	AATAACTGAA	960
	TTTCCTGATT	TAACTGGAAC	TGCAAACCTG	GAGAGTCTGA	CTTTAACTGG	AGCACAGATC	1020
	TCATCTCTTC	CTCAAACCGT	CTGCAATCAG	TTACCTAATC	TCCAAGTGCT	AGATCTGTCT	1080
	TACAACCTAT	TAGAAGATTT	ACCCAGTTTT	TCAGTCTGCC	AAAAGCTTCA	GAAAATTGAC	1140
15	CTAAGACATA	ATGAAATCTA	CGAAATTAAA	GTTGACACTT	TCCAGCAGTT	GCTTAGCCTC	1200
	CGATCGCTGA	ATTTGGCTTG	GAACAAAATT	GCTATTATTC	ACCCCAATGC	ATTTTCCACT	1260
	TTGCCATCCC	TAATAAAGCT	GGACCTATCG	TCCAACCTCC	TGTCGTCTTT	TCCTATAACT	1320
	GGGTTACATG	GTTTAACTCA	СТТААААТТА	ACAGGAAATC	ATGCCTTACA	GAGCTTGATA	1380
	TCATCTGAAA	ACTTTCCAGA	ACTCAAGGTT	ATAGAAATGC	CTTATGCTTA	CCAGTGCTGT	1440
20	GCATTTGGAG	TGTGTGAGAA	TGCCTATAAG	ATTTCTAATC	AATGGAATAA	AGGTGACAAC	1500
	AGCAGTATGG	ACGACCTTCA	TAAGAAAGAT	GCTGGAATGT	TTCAGGCTCA	AGATGAACGT	1560
	GACCTTGAAG	ATTTCCTGCT	TGACTTTGAG	GAAGACCTGA	AAGCCCTTCA	TTCAGTGCAG	1620
	TGTTCACCTT	CCCCAGGCCC	CTTCAAACCC	TGTGAACACC	TGCTTGATGG	CTGGCTGATC	1680
	AGAATTGGAG	TGTGGACCAT	AGCAGTTCTG	GCACTTACTT	GTAATGCTTT	GGTGACTTCA	1740
25	ACAGTTTTCA	GATCCCCTCT	GTACATTTCC	CCCATTAAAC	TGTTAATTGG	GGTCATCGCA	1800
	GCAGTGAACA	TGCTCACGGG	AGTCTCCAGT	GCCGTGCTGG	CTGGTGTGGA	TGCGTTCACT	1860
•	TTTGGCAGCT	TTGCACGACA	TGGTGCCTGG	TGGGAGAATG	GGGTTGGTTG	CCATGTCATT	1920
	GGTTTTTTGT	CCATTTTTGC	TTCAGAATCA	TCTGTTTTCC	TGCTTACTCT	GGCAGCCCTG	1980

GAGCGTGGGT TCTCTGTGAA ATATTCTGCA AAATTTGAAA CGAAAGCTCC ATTTTCTAGC 2040

	CTGAAAGTAA TCATTTTGCT CTGTGCCCTG CTGGCCTTGA CCATGGCCGC AGTTCCCCTG	2100
	CTGGGTGGCA GCAAGTATGG CGCCTCCCCT CTCTGCCTGC CTTTGCCTTT TGGGGAGCCC	2160
	AGCACCATGG GCTACATGGT CGCTCTCATC TTGCTCAATT CCCTTTGCTT CCTCATGATG	2220
. 5	ACCATTGCCT ACACCAAGCT CTACTGCAAT TTGGACAAGG GAGACCTGGA GAATATTTGG	2280
	GACTGCTCTA TGAAAAAACA CATTGCCCTG TTGCTCTTCA CCAACTGCAT CCTAAACTGC	2340
	CCTGTGGCTT TCTTGTCCTT CTCCTCTTTA ATAAACCTTA CATTTATCAG TCCTGAAGTA	2400
	ATTAAGTTTA TCCTTCTGGT GGTAGTCCCA CTTCCTGCAT GTCTCAATCC CCTTCTCTAC	2460
	ATCTTGTTCA ATCCTCACTT TAAGGAGGAT CTGGTGAGCC TGAGAAAGCA AACCTACGTC	2520
10	TGGACAAGAT CAAAACACCC AAGCTTGATG TCAATTAACT CTGATGATGT CGAAAAACAG	2580
	TCCTGTGACT CAACTCAAGC CTTGGTAACC TTTACCAGCT CCAGCATCAC TTATGACCTG	2640
	CCTCCCAGTT CCGTGCCATC ACCAGCTTAT CCAGTGACTG AGAGCTGCCA TCTTTCCTCT	2700
	GTGGCATTTG TCCCATGTCT CTAA	2724
	(279) INFORMATION FOR SEQ ID NO:278:	
15	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 907 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: (D) TOPOLOGY: not relevant 	
20	(ii) MOLECULE TYPE: protein	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:278:	•
	Met Asp Thr Ser Arg Leu Gly Val Leu Leu Ser Leu Pro Val Leu Le 1 5 10 15	u
25.	Gln Leu Ala Thr Gly Gly Ser Ser Pro Arg Ser Gly Val Leu Leu Ar 20 25 30	g
	Gly Cys Pro Thr His Cys His Cys Glu Pro Asp Gly Arg Met Leu Le 35 40 45	u
	Arg Val Asp Cys Ser Asp Leu Gly Leu Ser Glu Leu Pro Ser Asn Le 50 55 60	u
30	Ser Val Phe Thr Ser Tyr Leu Asp Leu Ser Met Asn Asn Ile Ser Gl 65 70 75 80	

Leu Leu Pro Asn Pro Leu Pro Ser Leu Arg Phe Leu Glu Glu Leu Arg

85

מאוכרותיות אות החחוות וא.

•	Leu	Ala	Gly	Asn 100		Leu	Thr	Tyr	Ile 105		Lys	Gly	Ala	Phe 110		Gl
· .	Leu	Tyr	Ser 115		Lys	Val	Leu	Met 120	Leu	Gln	Asn	Asn	Gln 125		Arg	His
5	Val	Pro 130	Thr	Glu	Ala	Leu	Gln 135	Asn	Leu	Arg	Ser	Leu 140		Ser	Leu	Arg
	Leu 145	Asp	Ala	Asn	His	Ile 150	Ser	Tyr	Val	Pro	Pro 155	Ser	Cys	Phe	Ser	Gly 160
10	Leu	His	Ser	Leu	Arg 165	His	Leu	Trp	Leu	Asp 170	Asp	Asn	Ala	Leu	Thr 175	Glu
	Ile	Pro	Val	Gln 180	Ala	Phe	Arg	Ser	Leu 185	Ser	Ala	Leu	Gln	Ala 190	Met	Thr
	Leu	Ala	Leu 195	Asn	Lys	Ile	His	His 200	Ile	Pro	Asp	Tyr	Ala 205	Phe	Gly	Asn
15	Leu	Ser 210	Ser	Leu	Val	Val	Leu 215	His	Leu	His	Asn	Asn 220	Arg	Ile	His	Ser
	Leu 225	Gly	Lys	Lys	Cys	Phe 230	Asp	Gly	Leu	His	Ser 235	Leu	Glu	Thr	Leu	Asp 240
20	Leu	Asn	Tyr	Asn	Asn 245	Leu	Asp	Glu	Phe	Pro 250	Thr	Ala	Ile	Arg	Thr 255	Leu
	Ser	Asn	Leu	Lys 260	Glu	Leu	Gly	Phe	His 265	Ser	Asn	Asn	Ile	Arg 270	Ser	Ile
	Pro	Glu	Lys 275	Ala	Phe	Val	Gly	Asn 280	Pro	Ser	Leu	Ile	Thr 285	Ile	His	Phe
25	Tyr	Asp 290	Asn	Pro	Ile	Gln	Phe 295	Val	Gly	Arg	Ser	Ala 300	Phe	Gln	His	Leu
	Pro 305	Glu	Leu	Arg	Thr	Leu 310	Thr	Leu	Asn	Gly	Ala 315	Ser	Gln	Ile	Thr	Glu 320
30	Phe	Pro	Asp	Leu	Thr 325	Gly	Thr	Ala	Asn	Leu 330	Glu	Ser	Leu	Thr	Leu 335	Thr
	Gly	Ala	Gln	Ile 340	Ser	Ser	Leu	Pro	Gln 345	Thr	Val	Cys	Asn	Gln 350	Leu	Pro
	Asn	Leu	Gln 355	Val	Leu	Asp	Leu	Ser 360	Tyr	Asn	Leu	Leu	Glu 365	Asp	Leu	Pro
3 5	Ser	Phe 370	Ser	Val	Cys	Gln	Lys 375	Leu	Gln	Lys	Ile	Asp 380	Leu	Arg	His	Asn
	Glu	Tle	ጥኒም	Glu	Tle	Laze	17a 1	Ner	Th~	Dhe	~1 -	~1 ~	T	*	0	T

	385	5				390) ,· .				395	i				400
	Arg	Ser	Leu	ı Asn	Leu 405	Ala	Trp	Asr	Lys	11e		Ile	: Ile	: His	Pro 415	Asn
5	Ala	Phe	e Ser	Thr 420	Leu	Pro	Ser	Leu	1le 425		Leu	Asp	Leu	Ser 430		Asn
	Leu	Lev	Ser 435	Ser	Phe	Pro	Ile	Thr 440	Gly	Leu	His	Gly	Leu 445		His	Leu
	Lys	Leu 450	Thr	Gly	Asn	His	Ala 455	Leu	Gln	Ser	Leu	Ile 460		Ser	Glu	Asn
10	Phe 465	Pro	Glu	Leu	Lys	Val 470	Ile	Glu	Met	Pro	Tyr 475	Ala	Tyr	Gln	Cys	Cys 480
	Ala	Phe	Gly	Val	Cys 485	Glu	Asn	Ala	Tyr	Lys 490	Ile	Ser	Asn	Gln	Trp 495	Asn
15	Lys	Gly	Asp	Asn 500	Ser	Ser	Met	Asp	Asp 505	Leu	His	Lys	Lys	Asp 510	Ala	Gly
	Met	Phe	Gln 515	Ala	Gln	Asp	Glu	Arg 520	Asp	Leu	Glu	Asp	Phe 525	Leu	Leu	Asp
	Phe	Glu 530	Glu	Asp	Leu	Lys	Ala 535	Leu	His	Ser	Val	Gln 540		Ser	Pro	Ser
20	Pro 545	Gly	Pro	Phe	Lys	Pro 550	Cys	Glu	His	Leu	Leu 555	Asp	Gly	Trp	Leu	Ile 560
	Arg	Ile	Gly	Val	Trp 565	Thr	Ile	Ala	Val	Leu 570	Ala	Leu	Thr	Cys ·	Asn 575	Ala
25	Leu	Val	Thr	Ser 580	Thr	Val	Phe	Arg	Ser 585	Pro	Leu	Tyr	Ile	Ser 590	Pro	Ile
	Lys	Leu	Leu 595	Ile	Gly	Val	Ile	Ala 600	Ala	Val	Asn	Met	Leu 605	Thr	Gly	Val
	Ser	Ser 610	Ala	Va1	Leu	Ala	Gly 615	Val	Asp	Ala	Phe	Thr 620	Phe	Gly	Ser	Phe
30	Ala 625	Arg	His	Gly	Ala	Trp 630	Trp	Glu	Asn	Gly	Val 635	Gly	Cys	His		Ile 640
	Gly	Phe	Leu	Ser	Ile 645	Phe	Ala	Ser		Ser 650	Ser	Val	Phe	Leu	Leu 655	Thr
35	Leu	Ala	Ala	Leu 660	Glu .	Arg	Gly	Phe	Ser 665	Val	Lys	Tyr		Ala 670	Lys	Phe
	Glu	Thr	Lys 675	Ala	Pro	Phe	Ser	Ser 680	Leu	Lys	Val	Ile	Ile 685	Leu	Leu	Cys

. WO 00/22129 PCT/US99/23938

235

		Ala	Leu 690	Leu	Ala	Leu	Thr	Met 695	Ala	Ala	Val	Pro	Leu 700	Leu	Gly	Gly	Ser
		Lys 705	Tyr	Gly	Ala	Ser	Pro 710	Leu	Cys	Leu	Pro	Leu 715	Pro	Phe	Gly	Glu	Pro 720
5		Ser	Thr	Met	Gly	Tyr 725	Met	Val	Ala	Leu	Ile 730	Leu	Leu	Asn	Ser	Leu 735	Cys
		Phe	Leu	Met	Met 740	Thr	Ile	Ala	Tyr	Thr 745	Lys	Leu	Tyr	Cys	Asn 750	Leu	Asp
10		Lys	Gly	Asp 755	Leu	Glu	Asn	Ile	Trp 760	Asp	Cys	Ser	Met	Lys 765	Lys	His	Ile
		Ala	Leu 770	Leu	Leu	Phe	Thr	Asn 775	Cys	Ile	Leu	Asn	Cys 780	Pro	Val	Ala	Phe
		Leu 785	Ser	Phe	Ser	Ser	Leu 790	Ile	Asn	Leu	Thr	Phe 795	Ile	Ser	Pro	Glu	Val 800
15		Ile	Lys	Phe	Ile	Leu 805	Leu	Val	Val	Val	Pro 810	Leu	Pro	Ala	Cys	Leu 815	Asn
			Leu		820					825					830		
20		Ser	Leu	Arg 835	Lys	Gln	Thr	Tyr	Val 840	Trp	Thr	Arg	Ser	Lys 845	His	Pro	Ser
		Leu	Met 850	Ser	Ile	Asn	Ser	Asp 855	Asp	Val	Glu	Lys	Gln 860	Ser	Cys	Asp	Ser
		Thr 865	Gln	Ala	Leu	Val	Thr 870	Phe	Thr	Ser	Ser	Ser 875	Ile	Thr	Tyr	Asp	Leu 880
25			Pro			885					890		Val	Thr	Glu	Ser 895	Cys
		His	Leu	Ser	Ser 900	Val	Ala	Phe	Val	Pro 905	Сув	Leu					
	(280)	INF	ORMA	TION	FOR	SEÇ	ID	NO:2	179:					•			
30		(i)	(B) (C)	LEN TYP STR	GTH: E: II ANDE	32 ucle DNES	base	pai cid ingl	.rs				• •				
35			.) MC					_			10:27	9:					

CATGCCAACC GGCCCGCGAG GCTGCTGCTG GT

(281) INFORMATION FOR SEQ ID NO:280:

5

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 32 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:280:

ACCAGCAGCA GCCTCGCGGG CCGGTTGGCA TG

Int. Ional Application No PCT/US 99/23938

		PCI/US 9	9/23938
A. CLASS IPC 7	C12N15/12 C07K14/72 G01N33	/50 G01N33/566	
According t	o International Patent Classification (IPC) or to both national class	ification and IPC	•
B. FIELDS	SEARCHED		
Minimum di IPC 7	ocumentation searched (classification system followed by classific C12N C07K G01N	cation symbols)	
Documenta	tion searched other than minimum documentation to the extent tha	at such documents are included in the fields a	earched
Electronic d	lata base consulted during the international search (name of data	base and, where practical, search terms use	d)
C. DOCUM	ENTS CONSIDERED TO BE RELEVANT		
Category °	Citation of document, with indication, where appropriate, of the	relevant passages	Relevant to claim No.
X	KJELSBERG M. A. ET AL.: "CONST. ACTIVATION OF THE ALPHA1B-ADRENG RECEPTOR BY ALL AMINO ACID SUBSTAT A SINGLE SITE" JOURNAL OF BIOLOGICAL CHEMISTRY vol. 267, no. 3, 25 January 1992 (1992-01-25), pa 1430-1433, XP002911764 ISSN: 0021-9258 the whole document	ERGIC FITUTIONS	1,2, 4-13, 15-33, 35-37,41
X Furth	er documents are listed in the continuation of box C.	Patent family members are listed	in annex.
"A" documer conside "E" earlier de filing da "L" documer which is citation "O" documer other m "P" documer later the	it which may throw doubts on priority claim(s) or cited to establish the publication date of another or other special reason (as specified) nt referring to an oral disclosure, use, exhibition or	"T" later document published after the inte or priority date and not in conflict with cited to understand the principle or the invention "X" document of particular relevance; the cannot be considered novel or cannot involve an inventive step when the document of particular relevance; the cannot be considered to involve an inventive step when the document is combined with one or mo ments, such combination being obviouin the art. "&" document member of the same patent for the company of the international and	the application but cory underlying the laimed invention be considered to current is taken alone laimed invention rentive step when the re other such docues to a person skilled
	March 2000	Date of mailing of the international sea	ion report
Name and ma	ailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl. Fax: (+31-70) 340-3016	Authorized officer Mand 1 , B	

1 101	/ U O :	77/	とンソンロ
-------	---------	-----	-------

SCHEER A. ET AL.: "CONSTITUTIVELY ACTIVE G PROTEIN-COUPLED RECEPTORS: POTENTIAL MECHANISMS OF RECEPTOR ACTIVATION" JOURNAL OF RECEPTOR AND SIGNAL TRANSDUCTION RESEARCH, vol. 17, no. 1/03, 1997, pages 57-73, XP000867531 ISSN: 1079-9893 the whole document WO 97 21731 A (NEW ENGLAND MEDICAL CENTER INC) 19 June 1997 (1997-06-19) the whole document, especially Fig. 2-3 WO 98 38217 A (HERRICK DAVIS KATHARINE; TEITLER MILT (US); EGAN CHRISTINA C (US)) 3 September 1998 (1998-09-03) the whole document, especially page 7, lines 24-27, and figure 4 PAUWELS P. J. ET AL.: "REVIEW: AMINO ACID DOMAINS INVOLVED IN CONSTITUTIVE		1,2,4,15-33,35-37,41 1,2,4,9-13,20-32,35-37,41 1,2,4-13,15-33,35-37,41
G PROTEIN-COUPLED RECEPTORS: POTENTIAL MECHANISMS OF RECEPTOR ACTIVATION" JOURNAL OF RECEPTOR AND SIGNAL TRANSDUCTION RESEARCH, vol. 17, no. 1/03, 1997, pages 57-73, XP000867531 ISSN: 1079-9893 the whole document WO 97 21731 A (NEW ENGLAND MEDICAL CENTER INC) 19 June 1997 (1997-06-19) the whole document, especially Fig. 2-3 WO 98 38217 A (HERRICK DAVIS KATHARINE ;TEITLER MILT (US); EGAN CHRISTINA C (US)) 3 September 1998 (1998-09-03) the whole document, especially page 7, lines 24-27, and figure 4 PAUWELS P. J. ET AL: "REVIEW: AMINO ACID		4-13, 15-33, 35-37,41 1,2,4, 9-13, 20-32, 35-37,41 1,2, 4-13, 15-33,
the whole document, especially Fig. 2-3 W0 98 38217 A (HERRICK DAVIS KATHARINE; TEITLER MILT (US); EGAN CHRISTINA C (US)) 3 September 1998 (1998-09-03) the whole document, especially page 7, lines 24-27, and figure 4 PAUWELS P. J. ET AL.: "REVIEW: AMINO ACID		9-13, 20-32, 35-37,41 1,2, 4-13, 15-33,
WO 98 38217 A (HERRICK DAVIS KATHARINE; TEITLER MILT (US); EGAN CHRISTINA C (US)) 3 September 1998 (1998-09-03) the whole document, especially page 7, lines 24-27, and figure 4 PAUWELS P. J. ET AL.: "REVIEW: AMINO ACID		4-13, 15-33,
PAUWELS P. J. ET AL.: "REVIEW AMINO ACID		
ACTIVATION OF G-PROTEIN-COUPLED RECEPTORS" MOLECULAR NEUROBIOLOGY, vol. 17, no. 1/03, 1998, pages 109-135, XP000866477 ISSN: 0893-7648 the whole document		1,2, 4-13, 15-33, 35-37,41
	ACTIVATION OF G-PROTEIN-COUPLED RECEPTORS" MOLECULAR NEUROBIOLOGY, vol. 17, no. 1/03, 1998, pages 109-135, XP000866477 ISSN: 0893-7648	ACTIVATION OF G-PROTEIN-COUPLED RECEPTORS" MOLECULAR NEUROBIOLOGY, vol. 17, no. 1/03, 1998, pages 109-135, XP000866477 ISSN: 0893-7648

_rnational application No.

PCT/US 99/23938

B x i Obs rvations wh r certain cla	ims were found unsear hable (Continuation of item 1 of first sheet)
This International Search Report has not been	established in respect of certain claims under Article 17(2)(a) for the following reasons:
Claims Nos.: because they relate to subject matter to	not required to be searched by this Authority, namely:
Z Claims Nos.: 34,38-40 because they relate to parts of the Inte an extent that no meaningful Internatio See FURTHER INFORMATION	ernational Application that do not comply with the prescribed requirements to such onal Search can be carried out, specifically:
Claims Nos.: because they are dependent claims an	nd are not drafted in accordance with the second and third sentences of Rule 6.4(a).
Box II Observations where unity of inv	rention is lacking (Continuation of item 2 of first sheet)
This International Searching Authority found mul	oltiple inventions in this international application, as follows:
As all required additional search fees we have a search fees which fees we have a search fees which fees we have a search fees which fees we have a search fees we have a search fees which fees which fees we have a search fees	vere timely paid by the applicant, this International Search Report covers all
searchable claims.	
2. As all searchable claims could be search of any additional fee.	ched without effort justifying an additional fee, this Authority did not invite payment
As only some of the required additional covers only those claims for which fees	search fees were timely paid by the applicant, this International Search Report were paid, specifically claims Nos.:
4. No required additional search fees were restricted to the invention first mentioner	e timely paid by the applicant. Consequently, this International Search Report is in the claims; it is covered by claims Nos.:
Remark on Protest	The additional search fees were accompanied by the applicant's protest. No protest accompanied the payment of additional search fees.

International Application No. PCT/US 99 /23938

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 34,38-40,42,43

Claims 34, 38-40, 42 and 43 refer to compounds with an agonistic effect on a GPCR without giving a true technical characterization. Moreover, no such specific compounds are defined in the application. In consequence, the scope of said claims is ambiguous and vague, and their subject-matter is not sufficiently disclosed and supported (Art. 5 and 6 PCT). No search can be carried out for such purely speculative claims whose wording is, in fact, a mere recitation of the results to be achieved.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

information on patent family members

Intl. ional Application No PCT/US 99/23938

Patent document cited in search report	rt ,	Publication date		Patent family member(s)	Publication date
WO 9721731	A	19-06-1997	US AU CA	5750353 A 1334397 A 2239293 A	12-05-1998 03-07-1997 19-06-1997
W0 9838217	Α	03-09-1998	EP AU	0869975 A 	14-10-1998

THIS PAGE BLANK (USPTO)